

Description

## 5       NOVEL PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

## BACKGROUND OF THE INVENTION

Within the field of genetic engineering, polynucleotides encoding proteins of interest have been identified and cloned by methods that require a detailed  
10 knowledge of the structure and/or function of the polynucleotide or the encoded protein. These methods include hybridization screening, polymerase chain reaction (PCR), and expression cloning.

With the more recent advent of large DNA sequence databases and the accompanying data analysis tools, identification of genes of interest is possible through  
15 the analysis of raw sequence data. Databases can be "mined" to locate sequences that resemble (are "homologous to") sequences of known function. Alignment of similar sequences can be used to place novel sequences within families of structurally similar sequences. These analytical tools can be combined with structural information obtained from, for example, X-ray crystallography to predict the higher order structure  
20 of a novel polypeptide. These analyses also facilitate prediction of polypeptide function. These recent technological advances have greatly increased the pace of gene discovery.

Genetic engineering has made available a number of genes and proteins of pharmaceutical or other economic importance. Such proteins include, for example,  
25 tissue plasminogen activator (t-PA) (U.S. Patent No. 4,766,075), coagulation factor VII (U.S. Patent No. 4,784,950), erythropoietin (U.S. Patent No. 4,703,008), platelet derived growth factor (U.S. Patent No. 4,889,919), and various industrial enzymes (e.g., U.S. Patents Nos. 5,965,384; 5,942,431; and 5,922,586).

Although estimates vary as to the amount of the human genome that has  
30 been identified to date, there remains a need in the art for further characterization of the human genome and the proteins encoded thereby. Previously unknown genes and proteins will be useful in the treatment and/or prevention of many human diseases, included diseases that have heretofore been refractory to treatment.

## 35       SUMMARY OF THE INVENTION

Within one aspect of the invention there is provided an isolated polypeptide comprising fifteen contiguous amino acid residues of a polypeptide as

shown in SEQ ID NO:M, wherein M is an even integer from 2 to 422. Within one embodiment, the isolated polypeptide is from 15 to 2235 amino acid residues in length. Within another embodiment, the at least fifteen contiguous amino acid residues of SEQ ID NO:M are operably linked via a peptide bond or polypeptide linker to a second

5 polypeptide selected from the group consisting of maltose binding protein, an immunoglobulin constant region, a polyhistidine tag, and a peptide as shown in SEQ ID NO:423. Within another embodiment, the polypeptide comprises at least 30 contiguous residues of SEQ ID NO:M. Within a further embodiment, the polypeptide comprises at least 47 contiguous residues of SEQ ID NO:M. Within additional embodiments, the

10 polypeptide is selected from the group consisting of polypeptides of SEQ ID NOS: 4, 6, 8, 10, 12, 16, 18, 24, 28, 42, 48, 54, 62, 66, 68, 70, 72, 82, 90, 92, 94, 96, 98, 102, 106, 108, 110, 112, 122, 124, 130, 134, 136, 138, 140, 156, 158, 162, 164, 166, 168, 174, 178, 180, 186, 202, 204, 206, 208, 210, 224, 230, 232, 234, 236, 240, 242, 250, 252, 254, 258, 262, 270, 272, 284, 286, 288, 294, 300, 302, 306, 310, 312, 314, 316, 322,

15 324, 328, 326, 336, 338, 342, 344, 348, 350, 366, 368, 374, 378, 386, 388, 396, 398, 402, 406, 408, 412, 416, and 420; the group consisting of polypeptides of SEQ ID NOS: 4, 6, 8, 12, 16, 18, 24, 28, 42, 48, 54, 62, 66, 68, 70, 72, 90, 92, 94, 96, 98, 102, 106, 108, 110, 112, 122, 124, 130, 134, 138, 140, 156, 158, 162, 164, 166, 168, 174, 178, 180, 202, 204, 206, 210, 224, 230, 234, 236, 240, 242, 252, 254, 258, 262, 270,

20 272, 284, 286, 288, 294, 300, 302, 306, 312, 314, 322, 324, 326, 336, 338, 342, 344, 348, 350, 366, 368, 374, 378, 386, 388, 396, 398, 402, 406, 408, 412, 416, and 420; the group consisting of polypeptides of SEQ ID NOS: 4, 6, 8, 12, 16, 18, 24, 28, 42, 48, 54, 66, 68, 70, 72, 90, 92, 94, 96, 98, 102, 106, 108, 110, 112, 122, 124, 130, 134, 138, 140, 156, 158, 162, 164, 166, 168, 174, 178, 180, 202, 204, 206, 210, 224, 230, 234,

25 236, 240, 242, 252, 254, 258, 262, 270, 272, 284, 286, 288, 294, 300, 302, 306, 312, 314, 322, 324, 326, 338, 342, 344, 348, 350, 366, 368, 374, 378, 386, 388, 396, 398, 402, 406, 408, 412, and 416; or the group consisting of polypeptides of SEQ ID NOS: 6, 8, 12, 18, 24, 42, 48, 54, 66, 68, 70, 72, 90, 92, 96, 98, 102, 106, 110, 122, 134, 138, 140, 156, 158, 162, 164, 168, 174, 178, 180, 204, 206, 210, 224, 230, 234, 236, 240,

30 242, 252, 254, 258, 270, 272, 284, 286, 288, 294, 300, 302, 306, 312, 314, 324, 326, 338, 342, 344, 348, 350, 366, 368, 374, 378, 386, 388, 396, 398, 402, 408, 412, and 416.

Within a second aspect of the invention there is provided an isolated, mature protein encoded by a polynucleotide sequence selected from the group

35 consisting of SEQ ID NO:N, wherein N is an odd integer from 1 to 421. Within certain embodiments, N is 3, 5, 7, 9, 11, 15, 17, 23, 27, 41, 47, 53, 61, 65, 67, 69, 71, 81, 89, 91, 93, 95, 97, 101, 105, 107, 109, 111, 121, 123, 129, 133, 135, 137, 139, 155,

157, 161, 163, 165, 167, 173, 177, 179, 185, 201, 203, 205, 207, 209, 223, 229, 231, 233, 235, 239, 241, 249, 251, 253, 257, 261, 269, 271, 283, 285, 287, 293, 299, 301, 305, 309, 311, 313, 315, 321, 323, 327, 325, 335, 337, 341, 343, 347, 349, 365, 367, 373, 377, 385, 387, 395, 397, 401, 405, 407, 411, 415, or 419; N is 3, 5, 7, 11, 15, 17, 23, 27, 41, 47, 53, 61, 65, 67, 69, 71, 89, 91, 93, 95, 97, 101, 105, 107, 109, 111, 121, 123, 129, 133, 137, 139, 155, 157, 161, 163, 165, 167, 173, 177, 179, 201, 203, 205, 209, 223, 229, 233, 235, 239, 241, 251, 253, 257, 261, 269, 271, 283, 285, 287, 293, 299, 301, 305, 311, 313, 321, 323, 325, 335, 337, 341, 343, 347, 349, 365, 367, 373, 377, 385, 387, 395, 397, 401, 405, 407, 411, 415, or 419; N is 3, 5, 7, 11, 15, 17, 23, 27, 41, 47, 53, 61, 65, 67, 69, 71, 89, 91, 93, 95, 97, 101, 105, 107, 109, 111, 121, 123, 129, 133, 137, 139, 155, 157, 161, 163, 165, 167, 173, 177, 179, 201, 203, 205, 209, 223, 229, 233, 235, 239, 241, 251, 253, 257, 261, 269, 271, 283, 285, 287, 293, 299, 301, 305, 311, 313, 321, 323, 325, 337, 341, 343, 347, 349, 365, 367, 373, 377, 385, 387, 395, 397, 401, 405, 407, 411, or 415; or N is 5, 7, 11, 17, 23, 41, 47, 53, 65, 67, 69, 71, 89, 91, 95, 97, 101, 105, 109, 121, 133, 137, 139, 155, 157, 161, 163, 167, 173, 177, 179, 203, 205, 209, 223, 229, 233, 235, 239, 241, 251, 253, 257, 269, 271, 283, 285, 287, 293, 299, 301, 305, 311, 313, 323, 325, 337, 341, 343, 347, 349, 365, 367, 373, 377, 385, 387, 395, 397, 401, 407, 411, or 415.

A third aspect of the invention provides isolated polynucleotides encoding the polypeptides disclosed above. Within certain embodiments of the invention the polynucleotides comprise a sequence of nucleotides as shown in SEQ ID NO:N, wherein N is an odd integer as defined above

Within a fourth aspect of the invention there is provided an expression vector comprising the following operably linked elements: a transcription promoter; a DNA segment encoding a polypeptide as shown in SEQ ID NO:M, wherein M is an even integer from 2 to 422; and a transcription terminator. Within certain embodiments, M is 4, 6, 8, 10, 12, 16, 18, 24, 28, 42, 48, 54, 62, 66, 68, 70, 72, 82, 90, 92, 94, 96, 98, 102, 106, 108, 110, 112, 122, 124, 130, 134, 136, 138, 140, 156, 158, 162, 164, 166, 168, 174, 178, 180, 186, 202, 204, 206, 208, 210, 224, 230, 232, 234, 236, 240, 242, 250, 252, 254, 258, 262, 270, 272, 284, 286, 288, 294, 300, 302, 306, 310, 312, 314, 316, 322, 324, 328, 326, 336, 338, 342, 344, 348, 350, 366, 368, 374, 378, 386, 388, 396, 398, 402, 406, 408, 412, 416, or 420; M is 4, 6, 8, 12, 16, 18, 24, 28, 42, 48, 54, 62, 66, 68, 70, 72, 90, 92, 94, 96, 98, 102, 106, 108, 110, 112, 122, 124, 130, 134, 138, 140, 156, 158, 162, 164, 166, 168, 174, 178, 180, 202, 204, 206, 210, 224, 230, 234, 236, 240, 242, 252, 254, 258, 262, 270, 272, 284, 286, 288, 294, 300, 302, 306, 312, 314, 322, 324, 326, 336, 338, 342, 344, 348, 350, 366, 368, 374, 378, 386, 388, 396, 398, 402, 406, 408, 412, 416, or 420; M is 4, 6, 8, 12, 16, 18, 24, 28, 42,

48, 54, 66, 68, 70, 72, 90, 92, 94, 96, 98, 102, 106, 108, 110, 112, 122, 124, 130, 134, 138, 140, 156, 158, 162, 164, 166, 168, 174, 178, 180, 202, 204, 206, 210, 224, 230, 234, 236, 240, 242, 252, 254, 258, 262, 270, 272, 284, 286, 288, 294, 300, 302, 306, 312, 314, 322, 324, 326, 338, 342, 344, 348, 350, 366, 368, 374, 378, 386, 388, 396, 398, 402, 406, 408, 412, or 416; or M is 6, 8, 12, 18, 24, 42, 48, 54, 66, 68, 70, 72, 90, 92, 96, 98, 102, 106, 110, 122, 134, 138, 140, 156, 158, 162, 164, 168, 174, 178, 180, 204, 206, 210, 224, 230, 234, 236, 240, 242, 252, 254, 258, 270, 272, 284, 286, 288, 294, 300, 302, 306, 312, 314, 324, 326, 338, 342, 344, 348, 350, 366, 368, 374, 378, 386, 388, 396, 398, 402, 408, 412, or 416.

10 A fifth aspect of the invention provides a cultured cell comprising the expression vector disclosed above. The cultured cell can be used, *inter alia*, within a method of producing a polypeptide, the method comprising (a) culturing the cell under conditions whereby the sequence of nucleotides is expressed, and (b) recovering the polypeptide. The invention also provides a polypeptide produced by this method.

15 Within a sixth aspect of the invention there is provided an isolated polynucleotide encoding a fusion protein, wherein the fusion protein comprises a secretory peptide selected from the group consisting of secretory peptides shown in SEQ ID NO:M, wherein M is an even integer as defined above, operably linked to a second polypeptide.

20 Within a seventh aspect of the invention there is provided an expression vector comprising the following operably linked elements: a transcription promoter; a DNA segment encoding a fusion protein as disclosed above; and a transcription terminator. The invention further provides a cultured cell comprising this expression vector, wherein the cell expresses the DNA segment and produces the encoded fusion protein. Also provided is a method of producing a protein comprising culturing the cell under conditions whereby the DNA segment is expressed, and recovering the second polypeptide. Within one embodiment the recovered second polypeptide is joined to a portion of a protein of SEQ ID NO: M, wherein M is an even integer as defined above.

25 Within a further aspect of the invention there is provided a computer-readable medium encoded with a data structure comprising SEQ ID NO:X, wherein X is an integer from 1 to 422.

30 Within an additional aspect of the invention there is provided an antibody that specifically binds to a protein selected from of the group consisting of SEQ ID NO:M, wherein M is an even integer as defined above.

35 These and other aspects of the invention will become evident upon reference to the following detailed description of the invention.

## DETAILED DESCRIPTION OF THE INVENTION

Prior to setting forth the invention in detail, it may be helpful to the understanding thereof to define the following terms:

The term "affinity tag" is used herein to denote a polypeptide segment that can be attached to a second polypeptide to provide for purification of the second polypeptide or provide sites for attachment of the second polypeptide to a substrate. In principal, any peptide or protein for which an antibody or other specific binding agent is available can be used as an affinity tag. Affinity tags include a poly-histidine tract, protein A (Nilsson et al., *EMBO J.* 4:1075, 1985; Nilsson et al., *Methods Enzymol.* 198:3, 1991), glutathione S transferase (Smith and Johnson, *Gene* 67:31, 1988), Glu-Glu affinity tag (Grussenmeyer et al., *Proc. Natl. Acad. Sci. USA* 82:7952-7954, 1985; see SEQ ID NO:423), substance P, Flag™ peptide (Hopp et al., *Biotechnology* 6:1204-1210, 1988), maltose binding protein (Kellerman and Ferenci, *Methods Enzymol.* 90:459-463, 1982; Guan et al., *Gene* 67:21-30, 1987), streptavidin binding peptide, thioredoxin, ubiquitin, cellulose binding protein, T7 polymerase, immunoglobulin constant domain, or other antigenic epitope or binding domain. See, in general, Ford et al., *Protein Expression and Purification* 2: 95-107, 1991. Affinity tags can be used individually or in combination. DNAs encoding affinity tags and otehr reagents are available from commercial suppliers (e.g., Pharmacia Biotech, Piscataway, NJ; Eastman Kodak, New Haven, CT; New England Biolabs, Beverly, MA).

The term "allelic variant" is used herein to denote any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term allelic variant is also used herein to denote a protein encoded by an allelic variant of a gene.

The terms "amino-terminal" and "carboxyl-terminal" are used herein to denote positions within polypeptides. Where the context allows, these terms are used with reference to a particular sequence or portion of a polypeptide to denote proximity or relative position. For example, a certain sequence positioned carboxyl-terminal to a reference sequence within a polypeptide is located proximal to the carboxyl terminus of the reference sequence, but is not necessarily at the carboxyl terminus of the complete polypeptide.

A "complement" of a polynucleotide molecule is a polynucleotide molecule having a complementary base sequence and reverse orientation as compared to a reference sequence. For example, the sequence 5' ATGCACGGG 3' is complementary to 5' CCCGTGCAT 3'.

“Corresponding to”, when used in reference to a nucleotide or amino acid sequence, indicates the position in a second sequence that aligns with the reference position when two sequences are optimally aligned.

The term “degenerate nucleotide sequence” denotes a sequence of  
5 nucleotides that includes one or more degenerate codons (as compared to a reference polynucleotide molecule that encodes a polypeptide). Degenerate codons encompass different triplets of nucleotides, but encode the same amino acid residue (i.e., GAU and GAC triplets each encode Asp).

The term “expression vector” is used to denote a DNA molecule, linear  
10 or circular, that comprises a segment encoding a polypeptide of interest operably linked to additional segments that provide for its transcription, wherein said segments are arranged in a way that does not exist naturally. Such additional segments include promoter and terminator sequences, and may also include one or more origins of replication, one or more selectable markers, an enhancer, a polyadenylation signal, etc.  
15 Expression vectors are generally derived from plasmid or viral DNA, or may contain elements of both.

The term “isolated”, when applied to a polynucleotide, denotes that the polynucleotide has been removed from its natural genetic milieu and is thus free of other extraneous or unwanted coding sequences, and is in a form suitable for use within  
20 genetically engineered protein production systems. Such isolated molecules are those that are separated from their natural environment and include cDNA and genomic clones. Isolated DNA molecules of the present invention are free of other genes with which they are ordinarily associated, but may include naturally occurring 5' and 3' untranslated regions such as promoters and terminators. The identification of  
25 associated regions will be evident to one of ordinary skill in the art (see for example, Dynan and Tijan, *Nature* 316:774-78, 1985).

An “isolated” polypeptide or protein is a polypeptide or protein that is found in a condition other than its native environment, such as apart from blood and animal tissue. In a preferred form, the isolated polypeptide or protein is substantially  
30 free of other polypeptides or proteins, particularly other polypeptides or proteins of animal origin. It is preferred to provide the polypeptides or proteins in a highly purified form, i.e. greater than 95% pure, more preferably greater than 99% pure. When used in this context, the term “isolated” does not exclude the presence of the same polypeptide or protein in alternative physical forms, such as dimers or alternatively glycosylated or  
35 derivatized forms.

A “mature protein” is a protein that is produced by cellular processing of a primary translation product of a DNA sequence. Such processing may include

removal of a secretory signal peptide, sometimes in combination with a propeptide. Mature sequences can be predicted from full-length sequences using methods known in the art for predicting cleavage sites. See, for example, von Heijne (*Nuc. Acids Res.* 14:4683, 1986). The sequence of a mature protein can be determined experimentally  
5 by expressing a DNA sequence of interest in a eukaryotic host cell and determining the amino acid sequence of the final product. For proteins lacking secretory peptides, the primary translation product will be the mature protein.

"Operably linked", when referring to DNA segments, indicates that the segments are arranged so that they function in concert for their intended purposes, e.g.,  
10 transcription initiates in the promoter and proceeds through the coding segment to the terminator. When referring to polypeptides, "operably linked" includes both covalently (e.g., by disulfide bonding) and non-covalently (e.g., by hydrogen bonding, hydrophobic interactions, or salt-bridge interactions) linked sequences, wherein the desired function(s) of the sequences are retained.

15 The term "ortholog" denotes a polypeptide or protein obtained from one species that is the functional counterpart of a polypeptide or protein from a different species. Sequence differences among orthologs are the result of speciation.

"Paralogs" are distinct but structurally related proteins made by an organism. Paralogs are believed to arise through gene duplication. For example,  $\alpha$ -  
20 globin,  $\beta$ -globin, and myoglobin are paralogs of each other.

A "polynucleotide" is a single- or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. Polynucleotides include RNA and DNA, and may be isolated from natural sources, synthesized *in vitro*, or prepared from a combination of natural and synthetic  
25 molecules. Sizes of polynucleotides are expressed as base pairs (abbreviated "bp"), nucleotides ("nt"), or kilobases ("kb"). Where the context allows, the latter two terms may describe polynucleotides that are single-stranded or double-stranded. When the term is applied to double-stranded molecules it is used to denote overall length and will be understood to be equivalent to the term "base pairs". It will be recognized by those  
30 skilled in the art that the two strands of a double-stranded polynucleotide may differ slightly in length and that the ends thereof may be staggered as a result of enzymatic cleavage; thus all nucleotides within a double-stranded polynucleotide molecule may not be paired. Such unpaired ends will in general not exceed 20 nt in length.

A "polypeptide" is a polymer of amino acid residues joined by peptide  
35 bonds, whether produced naturally or synthetically. Polypeptides of less than about 10 amino acid residues are commonly referred to as "peptides".

The term "promoter" is used herein for its art-recognized meaning to denote a portion of a gene containing DNA sequences that provide for the binding of RNA polymerase and initiation of transcription. Promoter sequences are commonly, but not always, found in the 5' non-coding regions of genes.

5 A "protein" is a macromolecule comprising one or more polypeptide chains. A protein may also comprise non-peptidic components, such as carbohydrate groups. Carbohydrates and other non-peptidic substituents may be added to a protein by the cell in which the protein is produced, and will vary with the type of cell. Proteins are defined herein in terms of their amino acid backbone structures;  
10 substituents such as carbohydrate groups are generally not specified, but may be present nonetheless.

A "secretory signal sequence" is a DNA sequence that encodes a polypeptide (a "secretory peptide") that, as a component of a larger polypeptide, directs the larger polypeptide through a secretory pathway of a cell in which it is synthesized.  
15 The larger polypeptide is commonly cleaved to remove the secretory peptide during transit through the secretory pathway.

The present invention is based in part upon the discovery of a group of novel, protein-encoding DNA molecules. These DNA molecules and the amino acid sequences that they encode are shown in SEQ ID NO:1 through SEQ ID NO:436.  
20 Sequence analysis predicts that each of the encoded proteins includes an amino-terminal secretory peptide. These secretory peptides are shown below in Table 1, wherein residue numbers are in reference to the indicated SEQ ID NO. As will be understood by those skilled in the art, the cleavage sites predicted by conventional models of secretory peptide cleavage (e.g., von Heijne, *Nuc. Acids Res.* 14:4683, 1986)  
25 are not always exact and may vary by as much as  $\pm 5$  residues. In addition, cleavage may occur at multiple sites within 5 residues of the indicated position. The mature form of any given protein may thus consist of a plurality of species differing at their amino termini.



Table 1

<u>Protein</u>	<u>SEQ ID NO:</u>	<u>Residues 1-</u>
AFP210015	2	14
AFP170681	4	26
AFP413680	6	28
AFP483037	8	14
AFP230872	10	27
AFP178828	12	14
AFP200134	14	23
AFP195796	16	22
AFP477303	18	18
AFP354334	20	25
AFP250287	22	17
AFP177000	24	26
AFP278176	26	21
AFP202885	28	18
AFP221312	30	23
AFP239757	32	22
AFP226311	34	20
AFP305901	36	20
AFP325549	38	20
AFP81988	40	14
AFP199200	42	20
AFP290395	44	23
AFP212675	46	20
AFP326051	48	17
AFP512441	50	18
AFP55098	52	15
AFP169796	54	21
AFP280706	56	25
AFP383165	58	23
AFP195467	60	26
AFP134225	62	22
AFP261193	64	28
AFP324422	66	28
AFP374312	68	28
AFP258118	70	24
AFP74517	72	25
AFP254653	74	18
AFP108666	76	21
AFP8766	78	15
AFP397185	80	20
AFP195042	82	21
AFP310695	84	26
AFP70022	86	19
AFP121670	88	22
AFP345861	90	15

AFP395942	92	16
AFP170291	94	21
AFP297548	96	22
AFP188135	98	28
AFP302388	100	19
AFP263430	102	17
AFP201273	104	18
AFP98983	106	25
AFP581958	108	20
AFP404202	110	19
AFP207203	112	15
AFP220790	114	19
AFP536326	116	23
AFP257473	118	22
AFP248380	120	16
AFP276202	122	20
AFP227568	124	23
AFP229039	126	20
AFP176297	128	17
AFP356885	130	17
AFP226938	132	16
AFP138504	134	29
AFP359196	136	24
AFP501809	138	27
AFP152733	140	15
AFP541394	142	23
AFP243183	144	20
AFP80739	146	18
AFP361806	148	26
AFP483930	150	21
AFP257336	152	25
AFP195800	154	23
AFP179530	156	19
AFP279267	158	14
AFP299766	160	29
AFP244615	162	16
AFP325761	164	22
AFP226024	166	22
AFP257094	168	27
AFP197103	170	27
AFP271855	172	17
AFP324816	174	29
AFP407963	176	25
AFP369635	178	17
AFP93743	180	28
AFP243230	182	15
AFP169316	184	21
AFP130852	186	15

AFP194191	188	22
AFP213472	190	21
AFP360430	192	22
AFP491309	194	21
AFP193428	196	23
AFP366534	198	22
AFP22706	200	27
AFP389012	202	14
AFP137186	204	24
AFP127023	206	21
AFP389687	208	16
AFP293220	210	25
AFP425535	212	25
AFP301494	214	25
AFP345421	216	19
AFP216667	218	26
AFP247951	220	29
AFP4464	222	22
AFP561930	224	28
AFP192851	226	22
AFP252759	228	20
AFP199044	230	20
AFP357958	232	28
AFP117501	234	15
AFP194554	236	23
AFP371069	238	23
AFP313600	240	19
AFP262739	242	18
AFP180730	244	27
AFP287227	246	28
AFP75785	248	26
AFP174843	250	15
AFP250422	252	15
AFP198645	254	17
AFP238111	256	16
AFP460626	258	24
AFP271081	260	14
AFP277752	262	16
AFP291338	264	15
AFP551038	266	22
AFP301579	268	20
AFP266188	270	16
AFP275580	272	28
AFP298054	274	21
AFP348226	276	23
AFP349106	278	23
AFP288248	280	15
AFP436476	282	19

AFP352125	284	14
AFP62060	286	25
AFP236718	288	21
AFP75775	290	25
AFP407487	292	23
AFP280451	294	27
AFP11675	296	29
AFP348656	298	16
AFP277451	300	19
AFP287436	302	14
AFP116043	304	28
AFP138740	306	26
AFP15192	308	17
AFP169968	310	27
AFP173341	312	23
AFP17588	314	23
AFP176427	316	20
AFP192633	318	14
AFP193013	320	15
AFP193881	322	16
AFP195562	324	16
AFP199922	326	18
AFP204736	328	17
AFP206179	330	27
AFP221877	332	23
AFP222758	334	26
AFP227032	336	24
AFP229269	338	27
AFP232213	340	25
AFP237679	342	21
AFP249599	344	28
AFP275215	346	21
AFP290397	348	26
AFP306591	350	18
AFP310297	352	20
AFP314720	354	19
AFP318671	356	29
AFP323575	358	21
AFP327160	360	20
AFP329002	362	29
AFP345415	364	24
AFP347179	366	24
AFP359138	368	23
AFP365372	370	17
AFP367284	372	23
AFP372822	374	26
AFP374595	376	29
AFP375952	378	25

AFP382913	380	17
AFP389184	382	23
AFP404208	384	20
AFP404279	386	29
AFP409112	388	26
AFP413111	390	19
AFP415635	392	15
AFP421092	394	17
AFP436666	396	25
AFP448623	398	19
AFP454192	400	20
AFP49026	402	28
AFP51688	404	28
AFP525341	406	16
AFP545268	408	15
AFP592620	410	22
AFP62197	412	23
AFP68229	414	25
AFP71288	416	15
AFP77851	418	27
AFP81957	420	15
AFP85168	422	27

A secretory peptide of a protein of the present invention can be used to direct the secretion of other proteins of interest from a host cell. Thus, the present invention provides, *inter alia*, fusions comprising such a secretory peptide of a protein disclosed herein operably linked to another protein of interest. The secretory peptide can be used to direct the secretion of other proteins of interest by joining a polynucleotide sequence encoding it, in the correct reading frame, to the 5' end of a sequence encoding the other protein of interest. Those skilled in the art will recognize that the resulting fused sequence may encode additional residues of a protein of the present invention at the amino terminus of the protein to be secreted. In the extreme case, the fusion may comprise an entire protein of the present invention fused to the amino terminus of a second protein, whereby secretion of the fusion protein is directed by the secretory peptide of the protein of the present invention. It will often be desirable to include a proteolytic cleavage site between the protein of the present invention (or portion thereof) and the other protein of interest. The joined polynucleotide sequences are then introduced into a host cell, which is cultured according to conventional methods. The protein of interest is then recovered from the culture media. Methods for introducing DNA into host cells, culturing the cells, and isolating recombinant proteins are known in the art. Representative methods are summarized below.

Within certain embodiments of the invention, the protein is selected from those listed in Table 2. Within related embodiments of the invention, the polynucleotide is selected from polynucleotides encoding the proteins listed in Table 2, i.e., for a protein of SEQ ID NO:M, the polynucleotide is SEQ ID NO:M-1.

5

Table 2

SEQ ID NO:	Protein	SEQ ID NO:	Protein
6	AFP413680	234	AFP117501
12	AFP178828	236	AFP194554
18	AFP477303	240	AFP313600
24	AFP177000	242	AFP262739
42	AFP199200	252	AFP250422
48	AFP326051	254	AFP198645
66	AFP324422	258	AFP460626
68	AFP374312	270	AFP266188
72	AFP74517	272	AFP275580
90	AFP345861	288	AFP236718
92	AFP395942	294	AFP280451
96	AFP297548	300	AFP277451
98	AFP188135	306	AFP138740
110	AFP404202	324	AFP195562
134	AFP138504	338	AFP229269
138	AFP501809	342	AFP237679
156	AFP179530	344	AFP249599
158	AFP279267	348	AFP290397
162	AFP244615	350	AFP306591
164	AFP325761	366	AFP347179
174	AFP324816	374	AFP372822
180	AFP93743	378	AFP375952
204	AFP137186	386	AFP404279
206	AFP127023	396	AFP436666
210	AFP293220	398	AFP448623
224	AFP561930	408	AFP545268
230	AFP199044	416	AFP71288

Higher order structures of the proteins of the present invention can be predicted by computer analysis using available software (e.g., the Insight II® viewer and homology modeling tools available from MSI, San Diego, CA; and King and Sternberg, *Protein Sci.* 5:2298-310, 1996). In addition, analytical algorithms permit the identification of homologies between newly discovered proteins and known proteins. Such homologies are indicative of related biological functions.

10

AFP254653 is 49% identical in sequence to human lysozyme C. Lysozyme C is a secreted bacteriolytic enzyme with similarity to the alpha-lactalbumins. Both are small alpha + beta proteins with six conserved cysteines forming a disulfide core comprising three disulfide bonds. AFP254653 may also  
5 exhibit bacteriolytic or other antimicrobial activity.

AFP581958 is 43% identical to wheat aluminum-induced protein, a member of the Bowman-Birk proteinase inhibitor family. All serine proteinases possess an exposed inhibitor loop that is stabilized by intermolecular interactions (usually disulfide bonds) between residues flanking the binding loop and the protein  
10 core. Interaction between inhibitor and enzyme produces a stable complex that disassociates very slowly, producing either an unaffected or a modified inhibitor that is cleaved at the scissile bond of the binding loop. AFP581958 may be a secreted serine proteinase.

AFP220790 is 42% identical to chicken lysozyme G, a bacteriolytic  
15 glycosyl hydrolase that hydrolyzes peptidoglycan homopolymers of the prokaryote cell walls. AFP220790 may thus be a secreted bacteriolytic enzyme, and may exhibit other antimicrobial activity.

AFP271855 is 37% identical to bovine granulocyte peptide A precursor (antimicrobial BGP-A). Bovine and murine granulocyte peptide A precursor (also  
20 called antimicrobial BGP-A) are disclosed in WIPO publication WO 97/29765. Bovine GP-A was isolated from a bone marrow library (WO 97/29765). GP-A exhibits activity against Gram-positive and Gram-negative bacteria, fungi and viruses. AFP271855 may exhibit antimicrobial (including one or more of anti-bacterial, anti-fungal, and anti-viral) activity.

AFP298054 is 24% identical to human T1/ST2 ligand. The T1 gene is  
25 also known as ST2, DER4, and Fit-1. It encodes a member of the interleukin-1 (IL-1) receptor family. It is transcribed in two forms, a soluble form and a membrane-bound form. The classical IL-1 ligands (IL-1 $\alpha$ , IL-1 $\beta$ , and IL-1ra) do not bind T1. A putative ligand for T1 was disclosed in 1996 (Gayle et al., *J. Biol. Chem.* 271:5784-5789, 1996).  
30 This protein binds T1 but is unable to initiate signal transduction by the membrane-bound form. The ligand is apparently a type I membrane protein. It has a predicted molecular weight (excluding the signal sequence and transmembrane domain) of about 22 kD, and has no sequence or hydrophobicity profile similarity to the beta-trefoil cytokines IL-1 or the FGFs. AFP298054 may be an antagonist that binds the receptor  
35 and regulates the activity of an as yet undiscovered IL-1 homolog.

Table 3 lists homologies between AFP sequences and sequences contained in the GenBank database, Derwent protein (PSP) or polynucleotide (PSN) databases, or Protein Identification Resource (PIR).

5

Table 3

Locus	Accession Number & Description
AFP130852	AE003823 (fly genomic)
AFP169968	AE003515 (fly genomic)
AFP174843	AF283518 ( <i>Mus musculus</i> elongation factor sec)
AFP176427	AE003808 (fly genomic)
AFP178828	PSN_V61483
AFP179530	AE003708 (fly genomic)
AFP188135	AE003677 (fly genomic)
AFP195042	PIR_T41241 (yeast oxysterol-binding protein family)
AFP198645	AE003718 (fly genomic)
AFP199200	AF113691 (human clone FLB4739 PRO1238 mRNA)
AFP204736	AC069237 (human chromosome 3 clone RP11-175M9)
AFP229269	AF247177 ( <i>Mus musculus</i> sphingosine-1-phosphate phosphohydrolase)
AFP230872	AF150741 ( <i>Rattus norvegicus</i> prolactin-like protein J mRNA)
AFP279267	AE003559 (fly genomic)
AFP347179	AE003499 (fly genomic) Z1041035F6P
AFP357958	AF283518 ( <i>Mus musculus</i> elongation factor sec mRNA)
AFP359196	AE003530 (fly genomic)
AFP374312	AE003538 (fly genomic)
AFP389687	AE003831 (fly genomic)
AFP395942	AB041564 (mouse brain cDNA; clone MNCb-0914)
AFP404202	AL137255 (human mRNA; cDNA DKFZp434B1813)
AFP413680	X14971 (mouse mRNA for alpha-adaptin, MMADAPA1)
AFP477303	AE003778 (fly genomic)
AFP62060	PSP_Y94938 (Human secreted protein clone ye78_1)
AFP71288	AL161655 (human chromosome 20 clone RP11-116E13)
AFP74517	PIR_T16263 ( <i>C. elegans</i> hypothetical protein F35D11.3)

Table 4 lists AFP proteins for which regions of identity have been found in the GenBank database.

Table 4

Locus	Accession Number & Description
AFP127023	SK000740 (human cDNA FLJ20733; clone HEP08550; by homology: molybdopterin cofactor sulfurase)
AFP134225	AB020970 (human mRNA; partial cds and 3'UTR; up-regulated by BCG-CWS)
AFP195562	AK000382 (human cDNA FLJ20375; clone HUV00942)



AFP199044	HSU80813 (human nucleoside diphosphate kinase homolog DR-nm23)
AFP227032	AK001848 (human cDNA FLJ10986; clone PLACE1001869; weakly similar to L-RIBULOKINASE; EC 2.7.1.16)
AFP237679	AB000465 (human mRNA; exon 1; 2; 3; 4; clone:RES4-24B; in genomic region of Huntington's disease locus)
AFP262739	AK000135 (human cDNA FLJ20128; clone COL06181)
AFP369635	PSN_Z24827 (Human secreted protein gene 17 clone HNF1Y77)
AFP81957	AF267730 (human 26S proteasome-associated UCH interacting protein 1; UIP1)
AFP93743	AK000066 (human cDNA FLJ20059; clone COL01349)

Table 5 lists AFP proteins for which longer regions of identity have been found in proteins contained in GenBank and other databases.

Table 5

Locus	Accession Number & Description
AFP117501	AK000505 (human cDNA FLJ20498; clone KAT08960)
AFP138740	HSM802370 (human mRNA; cDNA DKFZp434M1511)
AFP170291	AK000494 (human cDNA FLJ20487; clone KAT08245)
AFP170681	AK001698 (human cDNA FLJ10836; clone NT2RP4001228 close paralogue of human Kelch-like 1 protein (KLHL1) mRNA: AF252283)
AFP177000	AK000524 (human cDNA FLJ20517; clone KAT10235)
AFP193881	AK000382 (human cDNA FLJ20375; clone HUV00942)
AFP195796	AF251041 (human SGC32445 protein (SGC32445) mRNA; homology to PSP_W35393 Human TB2 gene product)
AFP202885	AB037808 (human mRNA for KIAA1387 protein)
AFP207203	AF250924 (human PNGase mRNA: peptide N-glycanase)
AFP226024	AK001952 (human cDNA FLJ11090; clone PLACE1005308)
AFP227568	AB019038 (human HMT-1 mRNA for beta-1,4 mannosyltransferase)
AFP244615	AK001009 (human cDNA FLJ10147; clone HEMBA1003369; weak homology: CENE_HUMAN CENTROMERIC PROTEIN E)
AFP250422	AF208849 (human BM-007 mRNA)
AFP266188	AK000272 (human cDNA FLJ20265; clone COLF9334; homology to major facilitator protein homolog, fission yeast: PIR_S62432)
AFP277451	AK001373 (human cDNA FLJ10511; clone NT2RP2000656)
AFP277752	AK000453 (human cDNA FLJ20446; clone KAT05231; weak homology to dinitrogenase reductase activating glycohydrolase (draG) <i>Archaeoglobus fulgidus</i> : PIR_C69465)
AFP280451	AL133355 (Human DNA sequence from clone RP11-541N10 on chromosome 10. Contains a novel gene and the 5' end of the gene for a novel protein; ortholog of mouse FISH protein)
AFP293220	AK001441 (human cDNA FLJ10579; clone NT2RP2003446)
AFP297548	AK000494 (human cDNA FLJ20487; clone KAT08245)
AFP306591	AL359700 (human chromosome 6 clone RP11-802L12)
AFP324816	AB032966 (human mRNA for KIAA1140 protein weak homology: Human O-linked GlcNAc transferase mRNA)

AFP356885	AK001544 (human cDNA FLJ10682; clone NT2RP3000072)
AFP389012	AK000428 (human cDNA FLJ20421; clone KAT02467; homologous to human bisphosphate 3'-nucleotidase mRNA: AF125042)
AFP436666	AK001608 (human cDNA FLJ10746; clone NT2RP3001679; likely human orthologue of <i>Rattus norvegicus</i> small rec (srec) mRNA: AF228917)
AFP501809	AK001963 (human cDNA FLJ11101; clone PLACE1005623)
AFP525341	AF189692 (human non-kinase Cdc42 effector protein SPEC2 mRNA)

A protein of the present invention can be prepared as a fusion protein by joining it to a second polypeptide or a plurality of additional polypeptides. Suitable second polypeptides include amino- or carboxyl-terminal extensions, such as linker peptides of up to about 20-25 residues and extensions that facilitate purification (affinity tags) as disclosed above. A protein of interest can be prepared as a fusion to a dimerizing protein as disclosed in U.S. Patents Nos. 5,155,027 and 5,567,584. Preferred dimerizing proteins in this regard include immunoglobulin constant region domains. Immunoglobulin-polypeptide fusions can be expressed in genetically engineered cells to produce a variety of multimeric analogs of a protein of interest. Fusion proteins can also comprise auxiliary domains that target the protein of interest to specific cells, tissues, or macromolecules (e.g., collagen). For example, a protein of interest can be targeted to a predetermined cell type by fusing it to a ligand that specifically binds to a receptor on the surface of a target cell. In this way, proteins can be targeted for therapeutic or diagnostic purposes. A protein can be fused to two or more moieties, such as an affinity tag for purification and a targeting domain. Protein fusions can also comprise one or more cleavage sites, particularly between domains. See, Tuan et al., *Connective Tissue Research* 34:1-9, 1996. Proteins of the present invention can also be used as targetting moieties within fusion proteins comprising, for example, cytokines, cytotoxins, or other biologically active polypeptide moieties.

Protein fusions of the present invention will usually contain not more than about 1,200 amino acid residues joined to the AFP protein. For example, an AFP protein can be fused to *E. coli*  $\beta$ -galactosidase (1,021 residues; see Casadaban et al., *J. Bacteriol.* 143:971-980, 1980), a 10-residue spacer, and a 4-residue factor Xa cleavage site. Such a protein comprising, for example, AFP345421 (SEQ ID NO:216), contains 2235 amino acid residues. In a second example, an AFP protein can be fused to maltose binding protein (approximately 370 residues), a 4-residue cleavage site, and a 6-residue polyhistidine tag.

As disclosed above, the proteins of the present invention or portions thereof can also be used to direct the secretion of a second protein. When such fusions

are designed so that the secreted protein retains a portion of the protein of the present invention, the fusion protein can be purified by means that exploit the properties of the protein of the present invention. Typical of such methods is immunoaffinity chromatography using an antibody directed against a protein of the present invention.

- 5 When such a fusion is engineered to contain a cleavage site at the fusion point, the fusion can be cleaved and the protein of interest recovered free of extraneous sequence.

The present invention also provides polynucleotide molecules, including DNA and RNA molecules, that encode the proteins disclosed above. Those skilled in the art will readily recognize that, in view of the degeneracy of the genetic code, considerable sequence variation is possible among these polynucleotide molecules. The amino acid sequence information provided herein can be used by one of ordinary skill in the art to generate degenerate sequences comprising all nucleotide sequences encoding a particular polypeptide. Table 6 sets forth the one-letter codes used to denote degenerate nucleotide positions. "Resolutions" are the nucleotides denoted by a code letter. "Complement" indicates the code for the complementary nucleotide(s). For example, the code Y denotes either C or T, and its complement R denotes A or G, A being complementary to T, and G being complementary to C.

TABLE 6

20

Nucleotide	Resolutions	Complement	Resolutions
A	A	T	T
C	C	G	G
G	G	C	C
T	T	A	A
R	A G	Y	C T
Y	C T	R	A G
M	A C	K	G T
K	G T	M	A C
S	C G	S	C G
W	A T	W	A T
H	A C T	D	A G T
B	C G T	V	A C G
V	A C G	B	C G T
D	A G T	H	A C T
N	A C G T	N	A C G T

Degenerate codons encompassing all possible codons for a given amino acid are set forth in Table 7, below.

TABLE 7

Amino Acid	One-Letter Code	Codons	Degenerate Codon
Cys	C	TGC TGT	TGY
Ser	S	AGC AGT TCA TCC TCG TCT	WSN
Thr	T	ACA ACC ACG ACT	CAN
Pro	P	CCA CCC CCG CCT	CCN
Ala	A	GCA GCC GCG GCT	GCN
Gly	G	GGA GGC GGG GGT	GGN
Asn	N	AAC AAT	AAY
Asp	D	GAC GAT	GAY
Glu	E	GAA GAG	GAR
Gln	Q	CAA CAG	CAR
His	H	CAC CAT	CAY
Arg	R	AGA AGG CGA CGC CGG CGT	MGN
Lys	K	AAA AAG	AAR
Met	M	ATG	ATG
Ile	I	ATA ATC ATT	ATH
Leu	L	CTA CTC CTG CTT TTA TTG	YTN
Val	V	GTA GTC GTG GTT	GTN
Phe	F	TTC TTT	TTY
Tyr	Y	TAC TAT	TAY
Trp	W	TGG	TGG
Ter	.	TAA TAG TGA	TRR
Asn Asp	B		RAY
Glu Gln	Z		SAR
Any	X		NNN
Gap	-	---	

5

One of ordinary skill in the art will appreciate that some ambiguity is introduced in determining a degenerate codon, representative of all possible codons encoding each amino acid. For example, the degenerate codon for serine (WSN) can, in some circumstances, encode arginine (AGR), and the degenerate codon for arginine (MGN) can, in some circumstances, encode serine (AGY). A similar relationship

10

exists between codons encoding phenylalanine and leucine. Thus, some polynucleotides encompassed by the degenerate sequences may encode variant amino acid sequences, but one of ordinary skill in the art can easily identify such variant sequences by reference to the amino acid sequences disclosed in the accompanying  
5 Sequence Listing.

Methods for preparing DNA and RNA are well known in the art. Complementary DNA (cDNA) clones are prepared from RNA that is isolated from a tissue or cell that produces large amounts of the cognate mRNA. Such tissues and cells are identified by methods commonly known in the art, such as Northern blotting  
10 (Thomas, *Proc. Natl. Acad. Sci. USA* 77:5201, 1980). Databases of expressed sequence tags (ESTs) can be analyzed to produce an "electronic Northern" wherein sequences are assigned to specific cell or tissue sources on the basis of their abundance within libraries. Table 8, below, shows the results of such an analysis when, as the minimum significant abundance, it was required that at least 10% of all sequences for a given  
15 protein were from a single source and at least five individual clones had been identified from that source. Sequences shown in the accompanying Sequence Listing but not listed in Table 8 were widely distributed among various tissues or were represented by few clones.

Table 8

AFP152733	K562 cells
AFP169796	T-cells
AFP173341	testis
AFP17588	fetal liver or spleen
AFP194554	fetal liver or spleen
AFP199922	testis
AFP229269	placenta
AFP237679	fetal liver or spleen
AFP257094	adult brain
AFP258118	epidermal breast keratinocytes
AFP263430	breast
AFP276202	infant brain
AFP287436	testis
AFP290397	testis
AFP306591	fetal heart
AFP325761	K562 cells
AFP352125	testis
AFP359138	infant brain
AFP369635	germinal center B-cells
AFP409112	kidney
AFP483037	neonatal keratinocytes
AFP49026	peripheral blood eosinophils of asthma patients
AFP545268	K562 cells
AFP561930	fetal liver or spleen
AFP62060	testis
AFP62197	pregnant uterus
AFP93743	germinal center B-cells
AFP98983	fetal heart

A panel of cDNAs from human tissues was screened for AFP expression using PCR. The panel was made from first strand cDNAs obtained from Clontech laboratories, Inc., Palo Alto, CA and contained 20 first-strand cDNA samples from the human tissues shown in Table 9. The panel was set up in a 96-well format that further included a human genomic DNA (obtained from Clontech Laboratories, Inc.) positive control sample and a water-only well as a negative control sample. Each well contained approximately 0.2-100 pg/ $\mu$ l of cDNA, diluted with water to 17.5 $\mu$ l. The

PCR reactions were set up by adding oligonucleotide primers, DNA polymerase (Ex Taq™; TAKARA Shuzo Co. Ltd. Biomedicals Group, Japan or Advantage™ 2 cDNA polymerase mix; Clontech Laboratories, Inc.) with the appropriate supplied buffer, dNTP mix (TAKARA Shuzo Co. Ltd.), and a density increasing agent and tracking dye  
5 (RediLoad; Research Genetics, Inc., Huntsville, AL) to each sample on the panel. The amplification was carried out as follows: incubation at 94°C for 2 minutes; 35 cycles of 94°C for 30 seconds, 60°C for 20 seconds, and 72°C for 30 seconds; followed by incubation at 72°C for 5 minutes. About 10 µl of the PCR reaction product was subjected to standard agarose gel electrophoresis using a 4% agarose gel.

Table 9

Protein	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
AFP117501	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	n
AFP127023	y	y	y	y	y	y	y	y	y	y	n	y	y	y	y	n	y	y	n	y	n	n
AFP137186	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP138504	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP138740	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	n
AFP177000	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP178828	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP179530	y	y	y	y	y	y	y	y	y	y	n	y	y	y	y	y	y	y	y	y	n	y
AFP188135	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	n
AFP194554	y	y	y	y	y	y	y	y	n	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP195562	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP198645	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
AFP199044	n	n	n	y	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
AFP199200	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP229269	y	y	y	y	y	y	y	y	y	y	y	y	n	y	y	y	y	y	y	y	n	y
AFP236718	n	y	y	y	y	y	y	n	n	y	n	n	n	y	n	n	n	y	y	y	n	y
AFP237679	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP244615	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	n
AFP249599	n	y	y	n	n	y	y	y	n	n	n	n	n	n	y	n	n	n	n	n	n	y
AFP250422	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP262739	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP266188	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP275580	y	y	y	y	y	y	y	n	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP277451	y	y	y	y	y	y	y	y	y	y	nd	y	y	y	y	y	y	y	y	y	n	n
AFP279267	y	y	y	y	y	y	y	y	n	y	y	y	n	y	y	y	y	y	y	y	n	n
AFP280451	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP290397	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP293220	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP297548	y	y	y	y	y	y	y	y	y	y	y	y	n	y	y	y	y	y	y	y	n	y
AFP306591	y	y	y	y	y	y	y	y	y	y	n	y	y	nd	y	nd	nd	y	y	y	n	y
AFP313600	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP324422	n	n	y	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	y	y	n	y



Table 9, continued

Protein	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
AFP324816	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP325761	y	y	y	y	y	y	y	y	y	y	y	y	n	y	y	y	y	y	y	y	n	n
AFP326051	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP345861	y	y	y	y	y	y	y	nd	nd	y	nd	y	nd	nd	y	y	y	y	y	y	n	n
AFP347179	y	y	y	y	y	y	y	y	y	y	y	y	n	y	y	y	n	y	y	y	n	y
AFP372822	y	y	y	y	y	y	y	y	n	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP374312	n	n	y	y	n	n	n	n	n	n	n	n	n	n	n	n	n	y	n	n	n	y
AFP375952	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	y
AFP395942	y	y	y	y	y	y	y	y	n	n	y	y	n	y	n	y	y	y	y	y	n	y
AFP404202	y	y	y	y	y	y	y	n	y	y	y	y	n	y	y	y	y	y	y	y	n	n
AFP404279	y	y	y	y	n	y	y	n	y	y	y	y	y	y	y	y	n	y	y	y	n	y
AFP413680	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	nd	y	y	y	n	y
AFP436666	y	y	y	y	y	y	y	y	n	y	y	y	n	y	y	y	y	y	y	y	n	n
AFP448623	y	n	y	n	n	y	n	n	n	n	n	n	n	n	y	n	n	y	y	n	n	y
AFP460626	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	nd	y	y	y	n	y
AFP477303	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP501809	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP545268	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y
AFP561930	y	y	y	y	y	y	y	nd	y	y	n	y	y	nd	y	y	y	y	nd	y	n	y
AFP71288	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	y
AFP74517	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	n
AFP93743	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	n	y

Tissues screened were: 1, brain; 2, heart; 3, kidney; 4, liver; 5, lung; 6, pancreas; 7, placenta; 8, skeletal muscle; 9, colon; 10, ovary; 11, peripheral blood leukocytes; 12, prostate; 13, small intestine; 14, spleen; 15, testis; 16, thymus; 17, bone marrow; 18, fetal liver; 19, lymph node; 20, tonsil; 21, H<sub>2</sub>O; 22, genomic DNA. Y=yes; n=no; nd=not determined.

Total RNA can be prepared using guanidine HCl extraction followed by isolation by centrifugation in a CsCl gradient (Chirgwin et al., *Biochemistry* 18:52-94, 1979). Poly (A)<sup>+</sup> RNA is prepared from total RNA using the method of Aviv and Leder (*Proc. Natl. Acad. Sci. USA* 69:1408-1412, 1972). Complementary DNA (cDNA) is prepared from poly(A)<sup>+</sup> RNA using known methods. In the alternative, genomic DNA can be isolated. For some applications (e.g., expression in transgenic animals) it may be preferable to use a genomic clone, or to modify a cDNA clone to include at least one genomic intron. Methods for identifying and isolating cDNA and genomic clones are well known and within the level of ordinary skill in the art, and include the use of the sequences disclosed herein, sequences complementary thereto, or parts thereof, for probing or priming a library. Such methods include, for example, hybridization or polymerase chain reaction ("PCR", Mullis, U.S. Patent 4,683,202). Expression libraries can be probed with antibodies to a protein of interest, receptor fragments, or other specific binding partners.

The polynucleotides of the present invention can also be prepared by automated synthesis. Synthesis of polynucleotides is within the level of ordinary skill in the art, and suitable equipment and reagents are available from commercial suppliers. See, in general, Glick and Pasternak, Molecular Biotechnology, Principles & Applications of Recombinant DNA, ASM Press, Washington, D.C., 1994; Itakura et al., *Ann. Rev. Biochem.* 53: 323-56, 1984; and Climie et al., *Proc. Natl. Acad. Sci. USA* 87:633-7, 1990.

The present invention further provides antisense polynucleotides that are complementary to a segment of a polynucleotide as set forth in one of SEQ ID NO:N, wherein N is an odd integer from 1 to 435. Such antisense polynucleotides are designed to bind to the corresponding mRNA and inhibit its translation. Antisense polynucleotides are used to inhibit gene expression in cell culture or in a patient, and can be used as probes or primers for research or diagnostic purposes.

Probes and primers of the present invention comprise a suitable fragment, and may comprise up to the complete sequence, of a polynucleotide as shown in SEQ ID NO:N or the complement thereof, wherein N is an odd integer from 1 to 421. Probes will generally be at least 20 nucleotides in length, although somewhat shorter probes (14-17 nucleotides) can be used. PCR primers are at least 5 nucleotides in length, preferably 15 or more nt, more preferably 20-30 nt. Shorter polynucleotide probes and primers are referred to in the art as "oligonucleotides," and can be DNA or RNA. Probes will generally comprise an oligonucleotide linked to a label, such as a radionuclide.

Probes and primers as disclosed herein can be used for cloning allelic, orthologous, and paralogous sequences. Allelic variants of the disclosed sequences can be cloned by probing cDNA or genomic libraries from different individuals according to standard procedures. Orthologous sequences can be cloned using information and compositions provided by the present invention in combination with conventional cloning techniques. For example, a cDNA can be cloned using mRNA obtained from a tissue or cell type that expresses the protein. Suitable sources of mRNA can be identified by probing Northern blots with probes designed from the sequences disclosed herein. A library is then prepared from mRNA of a positive tissue or cell line. A cDNA can then be isolated by a variety of methods, such as by probing with a complete or partial human cDNA or with one or more sets of degenerate probes based on the disclosed sequences. A cDNA can also be cloned by PCR using primers designed from the sequences disclosed herein. Within an additional method, the cDNA library can be used to transform or transfect host cells, and expression of the cDNA of interest can be detected with an antibody to the encoded protein. Similar techniques can also be applied to the isolation of genomic clones. Orthologous and paralogous sequences can be identified from libraries by probing blots at low stringency and washing the blots at successively higher stringency until background is suitably reduced.

Probes and primers disclosed herein can be used to clone 5' non-coding regions of a corresponding gene. In view of the tissue-specific expression observed for certain proteins of the invention (Tables 8 and 9), promoters of these genes are expected to provide tissue-specific expression. Such promoter elements can thus be used to direct the tissue-specific expression of heterologous genes in, for example, transgenic animals or patients treated with gene therapy. Cloning of 5' flanking sequences also facilitates production of a protein of interest by "gene activation" as disclosed in U.S. Patent No. 5,641,670. Briefly, expression of an endogenous gene in a cell is altered by introducing into its locus a DNA construct comprising at least a targeting sequence, a regulatory sequence, an exon, and an unpaired splice donor site. The targeting sequence is a 5' non-coding sequence that permits homologous recombination of the construct with the endogenous locus, whereby the sequences within the construct become operably linked with the endogenous coding sequence. In this way, an endogenous promoter can be replaced or supplemented with other regulatory sequences to provide enhanced, tissue-specific, or otherwise regulated expression.

The polynucleotides of the present invention further include polynucleotides encoding the fusion proteins, including signal peptide fusions, disclosed above.

The present invention further provides a computer-readable medium  
5 encoded with a data structure that provides at least one of SEQ ID NO:1 through SEQ ID NO:436. Suitable forms of computer-readable media include magnetic media and optically-readable media. Examples of magnetic media include a hard or fixed drive, a random access memory (RAM) chip, a floppy disk, digital linear tape (DLT), a disk cache, and a ZIP® disk. Optically readable media are exemplified by compact discs  
10 (e.g., CD-read only memory (ROM), CD-rewritable (RW), and CD-recordable), digital versatile/video discs (DVD) (e.g., DVD-ROM, DVD-RAM, and DVD+RW), and carrier waves.

The polypeptides of the present invention, including full-length proteins, biologically active fragments, immunogenic fragments, and fusion proteins,  
15 can be produced in genetically engineered host cells according to conventional techniques. Suitable host cells are those cell types that can be transformed or transfected with exogenous DNA and grown in culture, and include bacteria, fungal cells, and cultured higher eukaryotic cells. Eukaryotic cells, particularly cultured cells of multicellular organisms, are generally preferred for the production of proteins  
20 having higher eukaryotic-type post-translational modifications (e.g.,  $\gamma$ -carboxylation) and for making proteins, especially secretory proteins, for pharmaceutical use in humans. Techniques for manipulating cloned DNA molecules and introducing exogenous DNA into a variety of host cells are disclosed by Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory  
25 Press, Cold Spring Harbor, NY, 1989, and Ausubel et al., eds., *Current Protocols in Molecular Biology*, Green and Wiley and Sons, NY, 1993.

In general, a DNA sequence encoding a polypeptide of interest is operably linked to other genetic elements required for its expression, generally including a transcription promoter and terminator, within an expression vector. The  
30 vector will also commonly contain one or more selectable markers and one or more origins of replication, although those skilled in the art will recognize that within certain systems selectable markers can be provided on separate vectors, and replication of the exogenous DNA can be achieved through integration into the host cell genome. Selection of promoters, terminators, selectable markers, vectors and other elements is a  
35 matter of routine design within the level of ordinary skill in the art. Many such elements are described in the literature and are available through commercial suppliers.

To direct a polypeptide into the secretory pathway of a host cell, a secretory signal sequence (also known as a leader sequence, prepro sequence or pre sequence) is provided in the expression vector. The secretory signal sequence may be that of the protein of interest, or may be derived from another secreted protein (e.g., t-PA; see U.S. Patent No. 5,641,655) or synthesized *de novo*. The secretory signal sequence is operably linked to the DNA sequence encoding the protein of interest, i.e., the two sequences are joined in the correct reading frame and positioned to direct the newly synthesized protein into the secretory pathway of the host cell. Secretory signal sequences are commonly positioned 5' to the DNA sequence encoding the protein of interest, although certain secretory signal sequences may be positioned elsewhere in the DNA sequence of interest (see, e.g., Welch et al., U.S. Patent No. 5,037,743; Holland et al., U.S. Patent No. 5,143,830).

Cultured mammalian cells are suitable hosts for use within the present invention. Methods for introducing exogenous DNA into mammalian host cells include calcium phosphate-mediated transfection (Wigler et al., *Cell* 14:725, 1978; Corsaro and Pearson, *Somatic Cell Genetics* 7:603, 1981; Graham and Van der Eb, *Virology* 52:456, 1973), electroporation (Neumann et al., *EMBO J.* 1:841-845, 1982), DEAE-dextran mediated transfection (Ausubel et al., *ibid.*), and liposome-mediated transfection (Hawley-Nelson et al., *Focus* 15:73, 1993; Ciccarone et al., *Focus* 15:80, 1993). The production of recombinant polypeptides in cultured mammalian cells is disclosed by, for example, Levinson et al., U.S. Patent No. 4,713,339; Hagen et al., U.S. Patent No. 4,784,950; Palmiter et al., U.S. Patent No. 4,579,821; and Ringold, U.S. Patent No. 4,656,134. Suitable cultured mammalian cells include the COS-1 (ATCC No. CRL 1650), COS-7 (ATCC No. CRL 1651), BHK (ATCC No. CRL 1632), BHK 570 (ATCC No. CRL 10314), 293 (ATCC No. CRL 1573; Graham et al., *J. Gen. Virol.* 36:59-72, 1977) and Chinese hamster ovary (e.g. CHO-K1; ATCC No. CCL 61) cell lines. Additional suitable cell lines are known in the art and available from public depositories such as the American Type Culture Collection, Rockville, Maryland. In general, strong transcription promoters are preferred, such as promoters from SV-40 or cytomegalovirus. See, e.g., U.S. Patent No. 4,956,288. Other suitable promoters include those from metallothionein genes (U.S. Patent Nos. 4,579,821 and 4,601,978) and the adenovirus major late promoter. Within an alternative embodiment, adenovirus vectors can be employed. See, for example, Garnier et al., *Cytotechnol.* 15:145-55, 1994.

Drug selection is generally used to select for cultured mammalian cells into which foreign DNA has been inserted. Such cells are commonly referred to as "transfectants". Cells that have been cultured in the presence of the selective agent and

are able to pass the gene of interest to their progeny are referred to as "stable transfectants." An exemplary selectable marker is a gene encoding resistance to the antibiotic neomycin. Selection is carried out in the presence of a neomycin-type drug, such as G-418 or the like. Selection systems can also be used to increase the expression level of the gene of interest, a process referred to as "amplification." Amplification is carried out by culturing transfectants in the presence of a low level of the selective agent and then increasing the amount of selective agent to select for cells that produce high levels of the products of the introduced genes. An exemplary amplifiable selectable marker is dihydrofolate reductase, which confers resistance to methotrexate. Other drug resistance genes (e.g. hygromycin resistance, multi-drug resistance, puromycin acetyltransferase) can also be used.

Insect cells can be infected with recombinant baculovirus, commonly derived from *Autographa californica* nuclear polyhedrosis virus (AcNPV). See, King and Possee, The Baculovirus Expression System: A Laboratory Guide, London, Chapman & Hall; O'Reilly et al., Baculovirus Expression Vectors: A Laboratory Manual, New York, Oxford University Press., 1994; and Richardson, Ed., Baculovirus Expression Protocols. Methods in Molecular Biology, Humana Press, Totowa, NJ, 1995. Recombinant baculovirus can also be produced through the use of a transposon-based system described by Luckow et al. (*J. Virol.* 67:4566-4579, 1993). This system, which utilizes transfer vectors, is commercially available in kit form (Bac-to-Bac™ kit; Life Technologies, Rockville, MD). See also, Hill-Perkins and Possee, *J. Gen. Virol.* 71:971-976, 1990; Bonning et al., *J. Gen. Virol.* 75:1551-1556, 1994; and Chazenbalk and Rapoport, *J. Biol. Chem.* 270:1543-1549, 1995.

For protein production, the recombinant virus is used to infect host cells, typically a cell line derived from the fall armyworm, *Spodoptera frugiperda* (e.g., Sf9 or Sf21 cells) or *Trichoplusia ni* (e.g., High Five™ cells; Invitrogen, Carlsbad, CA). See, in general, Glick and Pasternak, Molecular Biotechnology: Principles and Applications of Recombinant DNA, ASM Press, Washington, D.C., 1994. See also, U.S. Patent No. 5,300,435. Serum-free media are used to grow and maintain the cells. Suitable media formulations are known in the art and can be obtained from commercial suppliers. The cells are grown up from an inoculation density of approximately  $2-5 \times 10^5$  cells to a density of  $1-2 \times 10^6$  cells, at which time a recombinant viral stock is added at a multiplicity of infection (MOI) of 0.1 to 10, more typically near 3. Procedures used are generally described in available laboratory manuals (e.g., King and Possee, *ibid.*; O'Reilly et al., *ibid.*; Richardson, *ibid.*). See also, Guarino et al., U.S. Patent No. 5,162,222 and WIPO publication WO 94/06463.

Fungal cells, including yeast cells, can also be used within the present invention. Yeast species of particular interest in this regard include *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Pichia methanolica*. Methods for transforming *S. cerevisiae* cells with exogenous DNA and producing recombinant polypeptides  
5 therefrom are disclosed by, for example, Kawasaki, U.S. Patent No. 4,599,311; Kawasaki et al., U.S. Patent No. 4,931,373; Brake, U.S. Patent No. 4,870,008; Welch et al., U.S. Patent No. 5,037,743; and Murray et al., U.S. Patent No. 4,845,075. Transformed cells are selected by phenotype determined by the selectable marker, commonly drug resistance or the ability to grow in the absence of a particular nutrient  
10 (e.g., leucine). A preferred vector system for use in *Saccharomyces cerevisiae* is the *POT1* vector system disclosed by Kawasaki et al. (U.S. Patent No. 4,931,373), which allows transformed cells to be selected by growth in glucose-containing media. Suitable promoters and terminators for use in yeast include those from glycolytic enzyme genes (see, e.g., Kawasaki, U.S. Patent No. 4,599,311; Kingsman et al., U.S.  
15 Patent No. 4,615,974; and Bitter, U.S. Patent No. 4,977,092) and alcohol dehydrogenase genes. See also U.S. Patents Nos. 4,990,446; 5,063,154; 5,139,936 and 4,661,454.

Transformation systems for other yeasts, including *Hansenula polymorpha*, *Schizosaccharomyces pombe*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Ustilago maydis*, *Pichia pastoris*, *Pichia methanolica*, *Pichia guilliermondii*  
20 and *Candida maltosa* are known in the art. See, for example, Gleeson et al., *J. Gen. Microbiol.* 132:3459-3465, 1986 and Cregg, U.S. Patent No. 4,882,279. *Aspergillus* cells may be utilized according to the methods of McKnight et al., U.S. Patent No. 4,935,349. Methods for transforming *Acremonium chrysogenum* are disclosed by  
25 Sumino et al., U.S. Patent No. 5,162,228. Methods for transforming *Neurospora* are disclosed by Lambowitz, U.S. Patent No. 4,486,533. Production of recombinant proteins in *Pichia methanolica* is disclosed in U.S. Patents No. 5,716,808, 5,736,383, 5,854,039, and 5,888,768; and WIPO publications WO 99/14347 and WO 99/14320.

Other higher eukaryotic cells, including plant cells and avian cells, can  
30 also be used as hosts according to methods commonly known in the art. For example, the use of *Agrobacterium rhizogenes* as a vector for expressing genes in plant cells has been reviewed by Sinkar et al., *J. Biosci. (Bangalore)* 11:47-58, 1987.

Prokaryotic host cells, including strains of the bacteria *Escherichia coli*, *Bacillus* and other genera are also useful host cells within the present invention.  
35 Techniques for transforming these hosts and expressing foreign DNA sequences cloned therein are well known in the art (see, e.g., Sambrook et al., *ibid.*). When expressing a polypeptide in bacteria such as *E. coli*, the polypeptide may be retained in the

cytoplasm, typically as insoluble granules, or may be directed to the periplasmic space by a bacterial secretion sequence. In the former case, the cells are lysed, and the granules are recovered and denatured using, for example, guanidine isothiocyanate or urea. The denatured polypeptide can then be refolded and dimerized by diluting the denaturant, such as by dialysis against a solution of urea and a combination of reduced and oxidized glutathione, followed by dialysis against a buffered saline solution. In the latter case, the polypeptide can be recovered from the periplasmic space in a soluble and functional form by disrupting the cells (by, for example, sonication or osmotic shock) to release the contents of the periplasmic space and recovering the protein, thereby obviating the need for denaturation and refolding.

Transformed or transfected host cells are cultured according to conventional procedures in a culture medium containing nutrients and other components required for the growth of the chosen host cells. A variety of suitable media, including defined media and complex media, are known in the art and generally include a carbon source, a nitrogen source, essential amino acids, vitamins and minerals. Media may also contain such components as growth factors or serum, as required. The growth medium will generally select for cells containing the exogenously added DNA by, for example, drug selection or deficiency in an essential nutrient which is complemented by the selectable marker carried on the expression vector or co-transfected into the host cell.

It is preferred to purify the polypeptides and proteins of the present invention to  $\geq 80\%$  purity, more preferably to  $\geq 90\%$  purity, even more preferably  $\geq 95\%$  purity, and particularly preferred is a pharmaceutically pure state, that is greater than 99.9% pure with respect to contaminating macromolecules, particularly other proteins and nucleic acids, and free of infectious and pyrogenic agents. Preferably, a purified polypeptide or protein is substantially free of other polypeptides or proteins, particularly those of animal origin.

Expressed recombinant proteins (including single polypeptide chains, chimeric polypeptides, and polypeptide multimers) are purified by conventional protein purification methods, typically by a combination of chromatographic techniques. See, in general, Affinity Chromatography: Principles & Methods, Pharmacia LKB Biotechnology, Uppsala, Sweden, 1988; and Scopes, Protein Purification: Principles and Practice, Springer-Verlag, New York, 1994. Proteins comprising a polyhistidine affinity tag (typically about 6 histidine residues) are purified by affinity chromatography on a nickel chelate resin. See, for example, Houchuli et al., *Bio/Technol.* 6: 1321-1325, 1988. Proteins comprising a glu-glu tag can be purified by immunoaffinity chromatography essentially as disclosed by Grussenmeyer et al., *ibid.*



Proteins comprising other affinity tags can be purified by appropriate affinity chromatography methods, which are known in the art.

Proteins of the present invention and fragments thereof can also be prepared through chemical synthesis according to methods known in the art, including  
5 exclusive solid phase synthesis, partial solid phase methods, fragment condensation or classical solution synthesis. See, for example, Merrifield, *J. Am. Chem. Soc.* 85:2149, 1963; Stewart et al., Solid Phase Peptide Synthesis (2nd edition), Pierce Chemical Co., Rockford, IL, 1984; Bayer and Rapp, *Chem. Pept. Prot.* 3:3, 1986; and Atherton et al., Solid Phase Peptide Synthesis: A Practical Approach, IRL Press, Oxford, 1989.

10 Using methods known in the art, the proteins of the present invention can be prepared in a variety of modified or derivatized forms. For example, the proteins can be prepared glycosylated or non-glycosylated; pegylated or non-pegylated; and may or may not include an initial methionine amino acid residue.

Biological activities of the proteins of the present invention can be  
15 measured *in vitro* using cultured cells or *in vivo* by administering molecules of the claimed invention to the appropriate animal model. Many such assays and models are known in the art. Guidance in initial assay selection is provided by structural predictions and sequence alignments. However, even if no functional prediction is made, the activity of a protein can be elucidated by known methods, including, for  
20 example, screening a variety of target cells for a biological response, other *in vitro* assays, expression in a host animal, or through the use of transgenic and/or "knockout" animals. Through the application of robotics, many *in vitro* assays can be adapted to rapid, high-throughput screening of a large number of samples. Target cells for use in activity assays include, without limitation, vascular cells (especially endothelial cells  
25 and smooth muscle cells), hematopoietic (myeloid and lymphoid) cells, liver cells (including hepatocytes, fenestrated endothelial cells, Kupffer cells, and Ito cells), fibroblasts (including human dermal fibroblasts and lung fibroblasts), neurite cells (including astrocytes, glial cells, dendritic cells, and PC-12 cells), fetal lung cells, articular synoviocytes, pericytes, chondrocytes, osteoblasts, adipocytes, and prostate  
30 epithelial cells. Endothelial cells and hematopoietic cells are derived from a common ancestral cell, the hemangioblast (Choi et al., *Development* 125:725-732, 1998).

Biological activity can be measured with a silicon-based biosensor microphysiometer that measures the extracellular acidification rate or proton excretion associated with receptor binding and subsequent physiologic cellular responses. An  
35 exemplary such device is the Cytosensor™ Microphysiometer manufactured by Molecular Devices, Sunnyvale, CA. A variety of cellular responses, such as cell proliferation, ion transport, energy production, inflammatory response, regulatory and

receptor activation, and the like, can be measured by this method. See, for example, McConnell et al., *Science* 257:1906-1912, 1992; Pitchford et al., *Meth. Enzymol.* 228:84-108, 1997; Arimilli et al., *J. Immunol. Meth.* 212:49-59, 1998; and Van Liefde et al., *Eur. J. Pharmacol.* 346:87-95, 1998. The microphysiometer can be used for  
5 assaying adherent or non-adherent eukaryotic or prokaryotic cells. By measuring extracellular acidification changes in cell media over time, the microphysiometer directly measures cellular responses to various stimuli, including agonistic and antagonistic stimuli. Preferably, the microphysiometer is used to measure responses of a eukaryotic cell known to be responsive to the protein of interest, compared to a  
10 control eukaryotic cell that does not respond to the protein of interest. Responsive eukaryotic cells comprise cells into which a receptor for the protein of interest has been transfected, as well as naturally responsive cells. Differences in the response of cells exposed to the protein of interest, relative to a control not so exposed, are a direct measurement of protein-modulated cellular responses. Such responses can be assayed  
15 under a variety of stimuli. The present invention thus provides methods of identifying agonists and antagonists of proteins of interest, comprising providing cells responsive to a selected protein, culturing a first portion of the cells in the absence of a test compound, culturing a second portion of the cells in the presence of a test compound, and detecting a change in a cellular response of the second portion of the cells as  
20 compared to the first portion of the cells. The change in cellular response is shown as a measurable change in extracellular acidification rate. Culturing a third portion of the cells in the presence of the protein of interest and the absence of a test compound provides a positive control and a control to compare the agonist activity of a test compound with that of the protein of interest. Antagonists can be identified by  
25 exposing the cells to the protein of interest in the presence and absence of the test compound, whereby a reduction in protein-stimulated activity is indicative of antagonist activity in the test compound.

Assays measuring cell proliferation or differentiation are well known in the art. For example, assays measuring proliferation include such assays as  
30 chemosensitivity to neutral red dye (Cavanaugh et al., *Investigational New Drugs* 8:347-354, 1990), incorporation of radiolabelled nucleotides (as disclosed by, e.g., Raines and Ross, *Methods Enzymol.* 109:749-773, 1985; Wahl et al., *Mol. Cell Biol.* 8:5016-5025, 1988; and Cook et al., *Analytical Biochem.* 179:1-7, 1989), incorporation of 5-bromo-2'-deoxyuridine (BrdU) in the DNA of proliferating cells (Porstmann et al.,  
35 *J. Immunol. Methods* 82:169-179, 1985), and use of tetrazolium salts (Mosmann, *J. Immunol. Methods* 65:55-63, 1983; Alley et al., *Cancer Res.* 48:589-601, 1988; Marshall et al., *Growth Reg.* 5:69-84, 1995; and Scudiero et al., *Cancer Res.* 48:4827-

4833, 1988). Differentiation can be assayed using suitable precursor cells that can be induced to differentiate into a more mature phenotype. Assays measuring differentiation include, for example, measuring cell-surface markers associated with stage-specific expression of a tissue, enzymatic activity, functional activity or morphological changes (Watt, *FASEB*, 5:281-284, 1991; Francis, *Differentiation* 57:63-75, 1994; Raes, *Adv. Anim. Cell Biol. Technol. Bioprocesses*, 161-171, 1989). Effects of a protein on tumor cell growth and metastasis can be analyzed using the Lewis lung carcinoma model, for example as described by Cao et al., *J. Exp. Med.* 182:2069-2077, 1995. Activity of a protein on cells of neural origin can be analyzed using assays that measure effects on neurite growth as disclosed below.

In vitro assays for pro- and anti-inflammatory activity are known in the art. Exemplary activity assays include mitogenesis assays in which IL-1 responsive cells (e.g., D10.N4.M cells) are incubated in the presence of IL-1 or a test protein for 72 hours at 37°C in a 5% CO<sub>2</sub> atmosphere. IL-2 (and optionally IL-4) is added to the culture medium to enhance sensitivity and specificity of the assay. <sup>3</sup>H-thymidine is then added, and incubation is continued for six hours. The amount of label incorporated is indicative of agonist activity. See, Hopkins and Humphreys, *J. Immunol. Methods* 120:271-276, 1989; Greenfeder et al., *J. Biol. Chem.* 270:22460-22466, 1995. Stimulation of cell proliferation can also be measured using thymocytes cultured in a test protein in combination with phytohemagglutinin. IL-1 is used as a control. Proliferation is detected as <sup>3</sup>H-thymidine incorporation or metabolic breakdown of (MTT) (Mosman, *ibid.*).

Protein activity may also be detected using assays designed to measure induction of one or more growth factors or other macromolecules. Preferred such assays include those for determining the presence of hepatocyte growth factor (HGF), epidermal growth factor (EGF), transforming growth factor alpha (TGFα), interleukin-6 (IL-6), VEGF, acidic fibroblast growth factor (aFGF), angiogenin, and other macromolecules produced by the liver. Suitable assays include mitogenesis assays using target cells responsive to the macromolecule of interest, receptor-binding assays, competition binding assays, immunological assays (e.g., ELISA), and other formats known in the art. Metalloprotease secretion is measured from treated primary human dermal fibroblasts, synoviocytes and chondrocytes. The relative levels of collagenase, gelatinase and stromalysin produced in response to culturing a target cell in the presence of a protein of interest is measured using zymogram gels (Loita and Stetler-Stevenson, *Cancer Biology* 1:96-106, 1990). Procollagen/collagen synthesis by dermal fibroblasts and chondrocytes in response to a test protein is measured using <sup>3</sup>H-proline incorporation into nascent secreted collagen. <sup>3</sup>H-labeled collagen is visualized by

SDS-PAGE followed by autoradiography (Unemori and Amento, *J. Biol. Chem.* 265: 10681-10685, 1990). Glycosaminoglycan (GAG) secretion from dermal fibroblasts and chondrocytes is measured using a 1,9-dimethylmethyle blue dye binding assay (Farndale et al., *Biochim. Biophys. Acta* 883:173-177, 1986). Collagen and GAG  
5 assays are also carried out in the presence of IL-1 $\beta$  or TGF- $\beta$  to examine the ability of a protein to modify the established responses to these cytokines.

Monocyte activation assays are carried out (1) to look for the ability of a protein of interest to further stimulate monocyte activation, and (2) to examine the ability of a protein of interest to modulate attachment-induced or endotoxin-induced  
10 monocyte activation (Fuhlbrigge et al., *J. Immunol.* 138: 3799-3802, 1987). IL-1 $\beta$  and TNF $\alpha$  levels produced in response to activation are measured by ELISA (Biosource, Inc. Camarillo, CA). Monocyte/macrophage cells, by virtue of CD14 (LPS receptor), are exquisitely sensitive to endotoxin, and proteins with moderate levels of endotoxin-like activity will activate these cells.

15 Other metabolic effects of proteins can be measured by culturing target cells in the presence and absence of a protein and observing changes in adipogenesis, gluconeogenesis, glycogenolysis, lipogenesis, glucose uptake, or the like. Suitable assays are known in the art.

Hematopoietic activity of proteins can be assayed on various  
20 hematopoietic cells in culture. Preferred assays include primary bone marrow colony assays and later stage lineage-restricted colony assays, which are known in the art (e.g., Holly et al., WIPO Publication WO 95/21920). Marrow cells plated on a suitable semi-solid medium (e.g., 50% methylcellulose containing 15% fetal bovine serum, 10% bovine serum albumin, and 0.6% PSN antibiotic mix) are incubated in the  
25 presence of test polypeptide, then examined microscopically for colony formation. Known hematopoietic factors are used as controls. Mitogenic activity of a protein of interest on hematopoietic cell lines can be measured as disclosed above.

Cell migration is assayed essentially as disclosed by Kähler et al. (*Arteriosclerosis, Thrombosis, and Vascular Biology* 17:932-939, 1997). A protein is  
30 considered to be chemotactic if it induces migration of cells from an area of low protein concentration to an area of high protein concentration. A typical assay is performed using modified Boyden chambers with a polystyrene membrane separating the two chambers (Transwell; Corning Costar Corp.). The test sample, diluted in medium containing 1% BSA, is added to the lower chamber of a 24-well plate  
35 containing Transwells. Cells are then placed on the Transwell insert that has been pretreated with 0.2% gelatin. Cell migration is measured after 4 hours of incubation at 37°C. Non-migrating cells are wiped off the top of the Transwell membrane, and cells

attached to the lower face of the membrane are fixed and stained with 0.1% crystal violet. Stained cells are then extracted with 10% acetic acid and absorbance is measured at 600 nm. Migration is then calculated from a standard calibration curve. Cell migration can also be measured using the matrigel method of Grant et al.  
5 ("Angiogenesis as a component of epithelial-mesenchymal interactions" in Goldberg and Rosen, *Epithelial-Mesenchymal Interaction in Cancer*, Birkhäuser Verlag, 1995, 235-248; Baatout, *Anticancer Research* 17:451-456, 1997).

Proteins can be assayed for the ability to modulate axon guidance and growth. Suitable assays that detect changes in neuron growth patterns include, for  
10 example, those disclosed in Hastings, WIPO Publication WO 97/29189 and Walter et al., *Development* 101:685-96, 1987. Assays to measure the effects on neuron growth are well known in the art. For example, the C assay (e.g., Raper and Kapfhammer, *Neuron* 4:21-9, 1990 and Luo et al., *Cell* 75:217-27, 1993) can be used to determine collapsing activity of a protein of interest on growing neurons. Other methods that can  
15 assess protein-induced inhibition of neurite extension or divert such extension are also known. See, Goodman, *Annu. Rev. Neurosci.* 19:341-77, 1996. Conditioned media from cells expressing a protein of interest, or aggregates of such cells, can be placed in a gel matrix near suitable neural cells, such as dorsal root ganglia (DRG) or sympathetic ganglia explants, which have been co-cultured with nerve growth factor.  
20 Compared to control cells, protein-induced changes in neuron growth can be measured (as disclosed by, for example, Messersmith et al., *Neuron* 14:949-59, 1995 and Puschel et al., *Neuron* 14:941-8, 1995). Neurite outgrowth can be measured using neuronal cell suspensions grown in the presence of molecules of the present invention. See, for example, O'Shea et al., *Neuron* 7:231-7, 1991 and DeFreitas et al., *Neuron* 15:333-43,  
25 1995.

Cell adhesion activity is assayed essentially as disclosed by LaFleur et al. (*J. Biol. Chem.* 272:32798-32803, 1997). Briefly, microtiter plates are coated with the test protein, non-specific sites are blocked with BSA, and cells (such as smooth muscle cells, leukocytes, or endothelial cells) are plated at a density of approximately  
30  $10^4 - 10^5$  cells/well. The wells are incubated at 37°C (typically for about 60 minutes), then non-adherent cells are removed by gentle washing. Adhered cells are quantitated by conventional methods (e.g., by staining with crystal violet, lysing the cells, and determining the optical density of the lysate). Control wells are coated with a known adhesive protein, such as fibronectin or vitronectin.

35 Assays for angiogenic activity are also known in the art. For example, the effect of a protein of interest on primordial endothelial cells in angiogenesis can be assayed in the chick chorioallantoic membrane angiogenesis assay (Leung, *Science*

246:1306-1309, 1989; Ferrara, *Ann. NY Acad. Sci.* 752:246-256, 1995). Briefly, a small window is cut into the shell of an eight-day old fertilized egg, and a test substance is applied to the chorioallantoic membrane. After 72 hours, the membrane is examined for neovascularization. Other suitable assays include microinjection of early stage quail (*Coturnix coturnix japonica*) embryos as disclosed by Drake et al. (*Proc. Natl. Acad. Sci. USA* 92:7657-7661, 1995); the rodent model of corneal neovascularization disclosed by Muthukkaruppan and Auerbach (*Science* 205:1416-1418, 1979), wherein a test substance is inserted into a pocket in the cornea of an inbred mouse; and the hamster cheek pouch assay (Höckel et al., *Arch. Surg.* 128:423-429, 1993). Induction of vascular permeability, which is indicative of angiogenic activity, is measured in assays designed to detect leakage of protein from the vasculature of a test animal (e.g., mouse or guinea pig) after administration of a test compound (Miles and Miles, *J. Physiol.* 118:228-257, 1952; Feng et al., *J. Exp. Med.* 183:1981-1986, 1996). *In vitro* assays for angiogenic activity include the tridimensional collagen gel matrix model (Pepper et al. *Biochem. Biophys. Res. Comm.* 189:824-831, 1992 and Ferrara et al., *Ann. NY Acad. Sci.* 732:246-256, 1995), which measures the formation of tube-like structures by microvascular endothelial cells; and matrigel models (Grant et al., "Angiogenesis as a component of epithelial-mesenchymal interactions" in Goldberg and Rosen, *Epithelial-Mesenchymal Interaction in Cancer*, Birkhäuser Verlag, 1995, 235-248; Baatout, *Anticancer Research* 17:451-456, 1997), which are used to determine effects on cell migration and tube formation by endothelial cells seeded in matrigel, a basement membrane extract enriched in laminin. It is preferred to carry out angiogenesis assays in the presence and absence of vascular endothelial growth factor (VEGF) to assess possible combinatorial effects. It is also preferred to use VEGF as a control within *in vivo* assays.

Receptor binding can be measured by the competition binding method of Labriola-Tompkins et al., *Proc. Natl. Acad. Sci. USA* 88:11182-11186, 1991. In an exemplary assay for IL-1 receptor binding, membranes prepared from EL-4 thymoma cells (Paganelli et al., *J. Immunol.* 138:2249-2253, 1987) are incubated in the presence of the test protein for 30 minutes at 37°C. Labeled IL-1 $\alpha$  or IL-1 $\beta$  is then added and the incubation is continued for 60 minutes. The assay is terminated by membrane filtration. The amount of bound label is determined by conventional means (e.g.,  $\gamma$  counter). In an alternative assay, the ability of a test protein to compete with labeled IL-1 for binding to cultured human dermal fibroblasts is measured according to the method of Dower et al. (*Nature* 324:266-268, 1986). Briefly, cells are incubated in a round-bottomed, 96-well plate in a suitable culture medium (e.g., RPMI 1640 containing 1% BSA, 0.1% Na azide, and 20 mM HEPES pH 7.4) at 8°C on a rocker

platform in the presence of labeled IL-1. Various concentrations of test protein are added. After the incubation (typically about two hours), cells are separated from unbound label by centrifuging 60- $\mu$ l aliquots through 200  $\mu$ l of phthalate oils in 400- $\mu$ l polyethylene centrifuge tubes and excising the tips of the tubes with a razor blade as disclosed by Segal and Hurwitz, *J. Immunol.* 118:1338-1347, 1977. Receptor binding assays for other cell types are known in the art. See, for example, Bowen-Pope and Ross, *Methods Enzymol.* 109:69-100, 1985.

Receptor binding can also be measured using immobilized receptors or ligand-binding receptor fragments. For example, an immobilized receptor can be exposed to its labeled ligand and unlabeled test protein, whereby a reduction in labeled ligand binding compared to a control is indicative of receptor-binding activity in the test protein. Within another format, a receptor or ligand-binding receptor fragment is immobilized on a biosensor (e.g., BIAcore™, Pharmacia Biosensor, Piscataway, NJ) and binding is determined. Antagonists of the native ligand will exhibit receptor binding but will exhibit essentially no activity in appropriate activity assays or will reduce the ligand-mediated response when combined with the native ligand. In view of the low level of receptor occupancy required to produce a response to some ligands (e.g., IL-1), a large excess of antagonist (typically a 10- to 1000-fold molar excess) may be necessary to neutralize ligand activity.

Receptor activation can be detected in target cells by: (1) measurement of adenylate cyclase activity (Salomon et al., *Anal. Biochem.* 58:541-48, 1974; Alvarez and Daniels, *Anal. Biochem.* 187:98-103, 1990); (2) measurement of change in intracellular cAMP levels using conventional radioimmunoassay methods (Steiner et al., *J. Biol. Chem.* 247:1106-13, 1972; Harper and Brooker, *J. Cyc. Nucl. Res.* 1:207-18, 1975); or (3) through use of a cAMP scintillation proximity assay (SPA) method (such as available from Amersham Corp., Arlington Heights, IL).

Proteins can be tested for serine protease activity or proteinase inhibitory activity using conventional assays. Substrate cleavage is conveniently assayed using a tetrapeptide that mimics the cleavage site of the natural substrate and which is linked, via a peptide bond, to a carboxyl-terminal para-nitro-anilide (pNA) group. The protease hydrolyzes the bond between the fourth amino acid residue and the pNA group, causing the pNA group to undergo a dramatic increase in absorbance at 405 nm. Suitable substrates can be synthesized according to known methods or obtained from commercial suppliers. Inhibitory activity is measured by adding a test sample to a reaction mixture containing enzyme and substrate, and comparing the observed enzyme activity to a control (without the test sample). A variety of such assays are known in the art, including assays measuring inhibition of trypsin,



chymotrypsin, plasmin, cathepsin G, and human leukocyte elastase. See, for example, Petersen et al., *Eur. J. Biochem.* 235:310-316, 1996. In a typical procedure, the inhibitory activity of a test compound is measured by incubating the test compound with the proteinase, then adding an appropriate substrate, typically a chromogenic peptide substrate. See, for example, Norris et al. (*Biol. Chem. Hoppe-Seyler* 371:37-42, 1990). Various concentrations of the inhibitor are incubated in the presence of trypsin, plasmin, and plasma kallikrein in a low-salt buffer at pH 7.4, 25°C. After 30 minutes, the residual enzymatic activity is measured by the addition of a chromogenic substrate (e.g., S2251 (D-Val-Leu-Lys-Nan) or S2302 (D-Pro-Phe-Arg-Nan), available from Kabi, Stockholm, Sweden) and a 30-minute incubation. Inhibition of enzyme activity is indicated by a decrease in absorbance at 405 nm or fluorescence Em at 460 nm. From the results, the apparent inhibition constant  $K_i$  is calculated. When a serine protease is prepared as an active precursor (e.g., comprising N-terminal residues 1-109 of SEQ ID NO:2), it is activated by cleavage with a suitable protease (e.g., furin (Steiner et al., *J. Biol. Chem.* 267:23435-23438, 1992)) prior to assay. Assays of this type are well known in the art. See, for example, Lottenberg et al., *Thrombosis Research* 28:313-332, 1982; Cho et al., *Biochem.* 23:644-650, 1984; Foster et al., *Biochem.* 26:7003-7011, 1987). The inhibition of coagulation factors (e.g., factor VIIa, factor Xa) can be measured using chromogenic substrates or in conventional coagulation assays (e.g., clotting time of normal human plasma; Dennis et al., *J. Biol. Chem.* 270:25411-25417, 1995).

Blood coagulation and chromogenic assays, which can be used to detect both procoagulant, anticoagulant, and thrombolytic activities, are known in the art. For example, pro- and anticoagulant activities can be measured in a one-stage clotting assay using platelet-poor or factor-deficient plasma (Levy and Edgington, *J. Exp. Med.* 151:1232-1243, 1980; Schwartz et al., *J. Clin. Invest.* 67:1650-1658, 1981). As disclosed by Anderson et al. (*Proc. Natl. Acad. Sci. USA* 96:11189-11193, 1999), the effect of a test compound on platelet activation can be determined by a change in turbidity, and the procoagulant activity of activated platelets can be determined in a phospholipid-dependent coagulation assay. Activation of thrombin can be determined by hydrolysis of peptide p-nitroanilide substrates as disclosed by Lottenberg et al. (*Thrombosis Res.* 28:313-332, 1982). Other procoagulant, anticoagulant, and thrombolytic activities can be measured using appropriate chromogenic substrates, a variety of which are available from commercial suppliers. See, for example, Kettner and Shaw, *Methods Enzymol.* 80:826-842, 1981.

Anti-microbial activity of proteins is evaluated by techniques that are known in the art. For example, anti-microbial activity can be assayed by evaluating the



sensitivity of microbial cell cultures to test agents and by evaluating the protective effect of test agents on infected mice. See, for example, Musiek et al., *Antimicrob. Agents Chemother.* 3:40, 1973. Antiviral activity can also be assessed by protection of mammalian cell cultures. Known techniques for evaluating anti-microbial activity include, for example, Barsum et al., *Eur. Respir. J.* 8:709-714, 1995; Sandovsky-Losica et al., *J. Med. Vet. Mycol (England)* 28:279-287, 1990; Mehentee et al., *J. Gen. Microbiol (England)* 135(:2181-2188, 1989; and Segal and Savage, *J. Med. Vet. Mycol.* 24:477-479, 1986. Assays specific for anti-viral activity include, for example, those described by Daher et al., *J. Virol.* 60:1068-1074, 1986.

10           The assays disclosed above can be modified by those skilled in the art to detect the presence of agonists and antagonists of a selected protein of interest.

          Expression of a polynucleotide encoding a protein of interest in animals provides models for further study of the biological effects of overproduction or inhibition of protein activity *in vivo*. Polynucleotides and antisense polynucleotides can be introduced into test animals, such as mice, using viral vectors or naked DNA, or transgenic animals can be produced.

          One *in vivo* approach for assaying proteins of the present invention utilizes viral delivery systems. Exemplary viruses for this purpose include adenovirus, herpesvirus, retroviruses, vaccinia virus, and adeno-associated virus (AAV). Adenovirus, a double-stranded DNA virus, is currently the best studied gene transfer vector for delivery of heterologous nucleic acids. For review, see Becker et al., *Meth. Cell Biol.* 43:161-89, 1994; and Douglas and Curiel, *Science & Medicine* 4:44-53, 1997. The adenovirus system offers several advantages. Adenovirus can (i) accommodate relatively large DNA inserts; (ii) be grown to high-titer; (iii) infect a broad range of mammalian cell types; and (iv) be used with many different promoters including ubiquitous, tissue specific, and regulatable promoters. Because adenoviruses are stable in the bloodstream, they can be administered by intravenous injection.

          By deleting portions of the adenovirus genome, larger inserts (up to 7 kb) of heterologous DNA can be accommodated. These inserts can be incorporated into the viral DNA by direct ligation or by homologous recombination with a co-transfected plasmid. In an exemplary system, the essential E1 gene is deleted from the viral vector, and the virus will not replicate unless the E1 gene is provided by the host cell (e.g., the human 293 cell line). When intravenously administered to intact animals, adenovirus primarily targets the liver. If the adenoviral delivery system has an E1 gene deletion, the virus cannot replicate in the host cells. However, the host's tissue (e.g., liver) will express and process (and, if a signal sequence is present, secrete) the

heterologous protein. Secreted proteins will enter the circulation in the highly vascularized liver, and effects on the infected animal can be determined.

An alternative method of gene delivery comprises removing cells from the body and introducing a vector into the cells as a naked DNA plasmid. The transformed cells are then re-implanted in the body. Naked DNA vectors are introduced into host cells by methods known in the art, including transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, use of a gene gun, or use of a DNA vector transporter. See, Wu et al., *J. Biol. Chem.* 263:14621-14624, 1988; Wu et al., *J. Biol. Chem.* 267:963-967, 1992; and Johnston and Tang, *Meth. Cell Biol.* 43:353-365, 1994.

Transgenic mice, engineered to express a gene encoding a protein of interest, and mice that exhibit a complete absence of gene function, referred to as "knockout mice" (Snouwaert et al., *Science* 257:1083, 1992), can also be generated (Lowell et al., *Nature* 366:740-742, 1993). These mice can be employed to study the gene of interest and the protein encoded thereby in an *in vivo* system. Transgenic mice are particularly useful for investigating the role of proteins in early development in that they allow the identification of developmental abnormalities or blocks resulting from the over- or underexpression of a specific factor. See also, Maisonpierre et al., *Science* 277:55-60, 1997 and Hanahan, *Science* 277:48-50, 1997. Preferred promoters for transgenic expression include promoters from metallothionein and albumin genes. As disclosed above, the human sequences provided herein can be used to clone orthologous polynucleotides, which may be preferred for use in generating transgenic and knockout animals.

Antisense methodology can be used to inhibit gene transcription to examine the effects of such inhibition *in vivo*. Polynucleotides that are complementary to a segment of a protein-encoding polynucleotide are designed to bind to the encoding mRNA and to inhibit translation of such mRNA. Such antisense oligonucleotides can also be used to inhibit expression of protein-encoding genes in cell culture.

Biological activities of test proteins can also be measured in animal models by administering the test protein, by itself or in combination with other agents, including other proteins. Using such models facilitates the assay of the test protein by itself or as an inhibitor or modulator of another agent, and also facilitates the measurement of combinatorial effects of bioactive compounds.

Anti-inflammatory activity can be tested in animal models of inflammatory disease. For example, animal models of psoriasis include the analysis of histological alterations in adult mouse tail epidermis (Hofbauer et al, *Brit. J. Dermatol.*

118:85-89, 1988; Bladon et al., *Arch Dermatol. Res.* 277:121-125, 1985). In this model, anti-psoriatic activity is indicated by the induction of a granular layer and orthokeratosis in areas of scale between the hinges of the tail epidermis. Typically, a topical ointment comprising a test compound is applied daily for seven consecutive  
5 days, then the animal is sacrificed, and tail skin is examined histologically. An additional model is provided by grafting psoriatic human skin to congenitally athymic (nude) mice (Krueger et al., *J. Invest. Dermatol.* 64:307-312, 1975). Such grafts have been shown to retain the characteristic histology for up to eleven weeks. As in the mouse tail model, the test composition is applied to the skin at predetermined intervals  
10 for a period of one to several weeks, at which time the animals are sacrificed and the skin grafts examined histologically. A third model has been disclosed by Fretland et al. (*Inflammation* 14:727-739, 1990). Briefly, inflammation is induced in guinea pig epidermis by topically applying phorbol ester (phorbol-12-myristate-13-acetate; PMA), typically at ca. 2 g/ml in acetone, to one ear and vehicle to the contralateral ear. Test  
15 compounds are applied concurrently with the PMA, or may be given orally. Histological analysis is performed at 96 hours after application of PMA. This model duplicates many symptoms of human psoriasis, including edema, inflammatory cell diapedesis and infiltration, high LTB<sub>4</sub> levels and epidermal proliferation.

Cerebral ischemia can be studied in a rat model as disclosed by Relton  
20 et al. (*ibid.*) and Loddick et al. (*ibid.*).

The effect of a test protein on primordial endothelial cells in angiogenesis can be assayed in the chick chorioallantoic membrane angiogenesis assay (Leung, *Science* 246:1306-1309, 1989; Ferrara, *Ann. NY Acad. Sci.* 752:246-256, 1995). Briefly, a small window is cut into the shell of an eight-day old fertilized egg,  
25 and a test substance is applied to the chorioallantoic membrane. After 72 hours, the membrane is examined for neovascularization. Embryo microinjection of early stage quail (*Coturnix coturnix japonica*) embryos can also be used (Drake et al., *Proc. Natl. Acad. Sci. USA* 92:7657-7661, 1995). Briefly, a solution containing the protein is injected into the interstitial space between the endoderm and the splanchnic mesoderm  
30 of early-stage embryos using a micropipette and micromanipulator system. After injection, embryos are placed ventral side down on a nutrient agar medium and incubated for 7 hours at 37°C in a humidified CO<sub>2</sub>/air mixture (10%/90%). Vascular development is assessed by microscopy of fixed, whole-mounted embryos and sections.

35 Stimulation of coronary collateral growth can be measured in known animal models, including a rabbit model of peripheral limb ischemia and hind limb ischemia and a pig model of chronic myocardial ischemia (Ferrara et al., *Endocrine*

*Reviews* 18:4-25, 1997). Test proteins are assayed in the presence and absence of VEGF and basic FGF to test for combinatorial effects. These models can be modified by the use of adenovirus or naked DNA for gene delivery as disclosed in more detail above, resulting in local expression of the test protein(s).

5           Angiogenic activity can also be tested in a rodent model of corneal neovascularization as disclosed by Muthukkaruppan and Auerbach, *Science* 205:1416-1418, 1979, wherein a test substance is inserted into a pocket in the cornea of an inbred mouse. For use in this assay, proteins are combined with a solid or semi-solid, biocompatible carrier, such as a polymer pellet. Angiogenesis is followed  
10   microscopically. Vascular growth into the corneal stroma can be detected in about 10 days.

          Angiogenic activity can also be tested in the hamster cheek pouch assay (Höckel et al., *Arch. Surg.* 128:423-429, 1993). A test substance is injected subcutaneously into the cheek pouch, and after five days the pouch is examined under  
15   low magnification to determine the extent of neovascularization. Tissue sections can also be examined histologically.

          Induction of vascular permeability is measured in assays designed to detect leakage of protein from the vasculature of a test animal (e.g., mouse or guinea pig) after administration of a test compound (Miles and Miles, *J. Physiol.* 118:228-257, 1952; Feng et al., *J. Exp. Med.* 183:1981-1986, 1996).  
20

          Wound-healing models include the linear skin incision model of Mustoe et al. (*Science* 237:1333, 1987). In a typical procedure, a 6-cm incision is made in the dorsal pelt of an adult rat, then closed with wound clips. Test substances and controls (in solution, gel, or powder form) are applied before primary closure. It is preferred to  
25   limit administration to a single application, although additional applications can be made on succeeding days by careful injection at several sites under the incision. Wound breaking strength is evaluated between 3 and 21 days post wounding. In a second model, multiple, small, full-thickness excisions are made on the ear of a rabbit. The cartilage in the ear splints the wound, removing the variable of wound contraction  
30   from the evaluation of closure. Experimental treatments and controls are applied. The geometry and anatomy of the wound site allow for reliable quantification of cell ingrowth and epithelial migration, as well as quantitative analysis of the biochemistry of the wounds (e.g., collagen content). See, Mustoe et al., *J. Clin. Invest.* 87:694, 1991. The rabbit ear model can be modified to create an ischemic wound environment,  
35   which more closely resembles the clinical situation (Ahn et al., *Ann. Plast. Surg.* 24:17, 1990). Within a third model, healing of partial-thickness skin wounds in pigs or guinea pigs is evaluated (LeGrand et al., *Growth Factors* 8:307, 1993). Experimental

treatments are applied daily on or under dressings. Seven days after wounding, granulation tissue thickness is determined. This model is preferred for dose-response studies, as it is more quantitative than other *in vivo* models of wound healing. A full thickness excision model can also be employed. Within this model, the epidermis and dermis are removed down to the panniculus carnosum in rodents or the subcutaneous fat in pigs. Experimental treatments are applied topically on or under a dressing, and can be applied daily if desired. The wound closes by a combination of contraction and cell ingrowth and proliferation. Measurable endpoints include time to wound closure, histologic score, and biochemical parameters of wound tissue. Impaired wound healing models are also known in the art (e.g., Cromack et al., *Surgery* 113:36, 1993; Pierce et al., *Proc. Natl. Acad. Sci. USA* 86:2229, 1989; Greenhalgh et al., *Amer. J. Pathol.* 136:1235, 1990). Delay or prolongation of the wound healing process can be induced pharmacologically by treatment with steroids, irradiation of the wound site, or by concomitant disease states (e.g., diabetes). Linear incisions or full-thickness excisions are most commonly used as the experimental wound. Endpoints are as disclosed above for each type of wound. Subcutaneous implants can be used to assess compounds acting in the early stages of wound healing (Broadley et al., *Lab. Invest.* 61:571, 1985; Sprugel et al., *Amer. J. Pathol.* 129: 601, 1987). Implants are prepared in a porous, relatively non-inflammatory container (e.g., polyethylene sponges or expanded polytetrafluoroethylene implants filled with bovine collagen) and placed subcutaneously in mice or rats. The interior of the implant is empty of cells, producing a "wound space" that is well-defined and separable from the preexisting tissue. This arrangement allows the assessment of cell influx and cell type as well as the measurement of vasculogenesis/angiogenesis and extracellular matrix production.

Inhibition of tumor metastasis can be assessed in mice into which cancerous cells or tumor tissue have been introduced by implantation or injection (e.g., Brown, *Advan. Enzyme Regul.* 35:293-301, 1995; Conway et al., *Clin. Exp. Metastasis* 14:115-124, 1996).

Effects on fibrinolysis can be measured in a rat model wherein the enzyme batroxobin and radiolabeled fibrinogen are administered to test animals. Inhibition of fibrinogen activation by a test compound is seen as a reduction in the circulating level of the label as compared to animals not receiving the test compound. See, Lenfors and Gustafsson, *Semin. Thromb. Hemost.* 22:335-342, 1996.

The invention further provides polypeptides that comprise an epitope-bearing portion of a protein as shown in SEQ ID NO:M, wherein M is an even integer from 2 to 436. An "epitope" is a region of a protein to which an antibody can bind. See, for example, Geysen et al., *Proc. Natl. Acad. Sci. USA* 81:3998-4002, 1984.

Epitopes can be linear or conformational, the latter being composed of discontinuous regions of the protein that form an epitope upon folding of the protein. Linear epitopes are generally at least 6 amino acid residues in length. Relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for example, Sutcliffe et al., *Science* 219:660-666, 1983. Antibodies that recognize short, linear epitopes are particularly useful in analytic and diagnostic applications that employ denatured protein, such as Western blotting (Tobin, *Proc. Natl. Acad. Sci. USA* 76:4350-4356, 1979). Antibodies to short peptides may also recognize proteins in native conformation and will thus be useful for monitoring protein expression and protein isolation, and in detecting proteins in solution, such as by ELISA or in immunoprecipitation studies.

Antigenic, epitope-bearing polypeptides of the present invention are useful for raising antibodies, including monoclonal antibodies, that specifically bind to the corresponding protein. Antigenic, epitope-bearing polypeptides contain a sequence of at least six, preferably at least nine, more preferably from 15 to about 30 contiguous amino acid residues of a protein. Within certain embodiments of the invention, the polypeptides comprise 40, 50, 100, or more contiguous residues of a protein as shown in SEQ ID NO:M, up to the entire predicted mature protein or the primary translation product. It is preferred that the amino acid sequence of the epitope-bearing polypeptide is selected to provide substantial solubility in aqueous solvents, that is the sequence includes relatively hydrophilic residues, and hydrophobic residues are substantially avoided. Table 10 lists preferred hexapeptides for use as antigens. Within Table 10, each the amino termini of the hexapeptides are specified. Those skilled in the art will recognize that longer polypeptides comprising these hexapeptides can also be used and will often be preferred.

Table 10

<u>Protein</u>		<u>Hexapeptide N-termini</u>			
AFP210015	389	405	97	388	359
AFP170681	51	334	113	49	140
AFP413680	221	207	220	206	198
AFP483037	219	218	82	216	215
AFP230872	189	188	73	156	68
AFP178828	211	210	209	208	207
AFP200134	150	149	146	132	145
AFP195796	99	97	111	208	240

AFP477303	64	126	63	54	112
AFP354334	269	268	267	266	265
AFP250287	34	33	48	2	143
AFP177000	133	132	104	37	68
AFP278176	234	145	284	91	291
AFP202885	134	244	170	133	243
AFP221312	31	29	28	51	43
AFP239757	329	200	556	107	328
AFP226311	293	74	250	86	184
AFP305901	340	194	451	192	120
AFP325549	293	74	250	86	184
AFP81988	151	167	147	165	173
AFP199200	150	149	148	92	147
AFP290395	31	29	28	329	326
AFP212675	67	66	65	204	396
AFP326051	49	56	23	78	95
AFP512441	94	93	41	39	38
AFP55098	140	34	139	120	32
AFP169796	177	173	156	32	155
AFP280706	33	54	32	31	53
AFP383165	25	82	52	24	178
AFP195467	113	112	71	2	80
AFP134225	114	280	113	455	417
AFP261193	120	66	65	85	119
AFP324422	147	145	66	65	85
AFP374312	125	124	79	123	77
AFP258118	64	63	116	115	62
AFP74517	1	72	124	123	22
AFP254653	134	36	62	14	23
AFP108666	79	76	74	49	48
AFP8766	140	34	139	120	298
AFP397185	265	35	264	34	48
AFP195042	192	535	191	259	533
AFP310695	49	75	190	5	94
AFP70022	38	64	179	83	37
AFP121670	184	183	121	118	182
AFP345861	151	89	75	135	149

AFP395942	60	14	59	13	21
AFP170291	144	72	56	55	63
AFP297548	145	73	57	56	64
AFP188135	152	148	158	147	144
AFP302388	478	431	416	414	429
AFP263430	92	23	64	91	110
AFP201273	373	384	163	372	44
AFP98983	3	2	35	34	32
AFP581958	71	66	80	26	25
AFP404202	1	31	115	30	92
AFP207203	427	258	204	426	48
AFP220790	139	92	51	187	91
AFP536326	87	146	105	73	103
AFP257473	270	205	203	245	244
AFP248380	283	62	54	272	100
AFP276202	50	48	35	46	33
AFP227568	199	23	238	363	224
AFP229039	226	91	116	161	225
AFP176297	261	382	183	119	182
AFP356885	622	45	525	175	466
AFP226938	118	108	117	79	107
AFP138504	77	255	75	254	292
AFP359196	4	76	3	2	37
AFP501809	141	139	9	169	2
AFP152733	258	204	48	47	257
AFP541394	31	29	28	235	232
AFP243183	272	110	106	3	2
AFP80739	398	397	224	223	155
AFP361806	4	78	139	3	76
AFP483930	107	124	123	88	45
AFP257336	124	42	122	182	158
AFP195800	40	39	65	38	96
AFP179530	57	251	249	315	55
AFP279267	106	62	216	187	59
AFP299766	127	168	165	29	126
AFP244615	171	196	326	255	179
AFP325761	138	137	2	144	109



AFP226024	79	317	159	140	45
AFP257094	71	116	115	3	144
AFP197103	200	198	215	195	177
AFP271855	92	44	42	18	27
AFP324816	9	252	120	8	63
AFP407963	202	201	156	200	155
AFP369635	98	398	255	97	254
AFP93743	4	254	3	294	293
AFP243230	28	129	128	127	44
AFP169316	294	170	293	36	157
AFP130852	82	59	117	145	66
AFP194191	363	112	271	69	267
AFP213472	103	102	69	2	37
AFP360430	177	75	183	74	130
AFP491309	107	106	69	2	37
AFP193428	129	87	343	60	128
AFP366534	72	4	2	59	39
AFP22706	229	227	65	64	188
AFP389012	216	27	289	34	17
AFP137186	2	1	182	216	43
AFP127023	86	56	131	178	55
AFP389687	57	56	117	370	369
AFP293220	186	194	105	146	182
AFP425535	264	181	163	370	149
AFP301494	159	4	2	84	25
AFP345421	500	592	639	652	849
AFP216667	92	435	329	422	47
AFP247951	27	34	33	25	94
AFP4464	365	363	362	55	209
AFP561930	108	107	104	52	66
AFP192851	300	276	299	298	496
AFP252759	311	310	64	21	157
AFP199044	143	2	209	206	125
AFP357958	167	338	165	324	362
AFP117501	135	87	362	86	418
AFP194554	318	170	54	105	169
AFP371069	332	1	283	365	279

AFP313600	341	340	240	48	176
AFP262739	25	24	142	23	207
AFP180730	58	37	30	27	36
AFP287227	596	592	591	374	525
AFP75785	128	127	136	99	71
AFP174843	152	323	150	309	347
AFP250422	100	140	99	138	182
AFP198645	145	144	143	64	56
AFP238111	123	50	20	137	35
AFP460626	153	151	71	150	70
AFP271081	68	112	39	202	67
AFP277752	109	106	220	238	92
AFP291338	347	342	97	362	339
AFP551038	134	131	186	130	173
AFP301579	105	153	130	152	67
AFP266188	121	235	61	180	120
AFP275580	193	77	192	2	148
AFP298054	148	234	146	233	144
AFP348226	148	103	85	309	59
AFP349106	208	118	117	207	116
AFP288248	376	342	340	339	312
AFP436476	18	39	139	38	99
AFP352125	53	59	163	142	104
AFP62060	247	187	73	426	72
AFP236718	100	99	249	248	184
AFP75775	201	90	239	173	199
AFP407487	148	103	85	59	58
AFP280451	141	294	6	209	139
AFP11675	58	56	90	64	89
AFP348656	160	159	158	103	149
AFP277451	118	2	1	146	241
AFP287436	53	59	223	142	104
AFP116043	212	239	138	186	183
AFP138740	264	263	31	72	232
AFP15192	47	46	216	85	212
AFP169968	64	117	63	2	81
AFP173341	65	64	102	101	100

AFP17588	43	42	2	41	1
AFP176427	311	290	308	155	288
AFP192633	58	56	162	349	44
AFP193013	47	90	87	46	68
AFP193881	274	295	402	273	292
AFP195562	274	295	339	473	273
AFP199922	57	55	74	180	50
AFP204736	89	58	43	28	23
AFP206179	74	80	73	71	70
AFP221877	32	31	30	50	75
AFP222758	44	43	75	42	19
AFP227032	47	55	46	65	54
AFP229269	147	127	146	63	60
AFP232213	44	41	28	27	40
AFP237679	2	1	34	58	55
AFP249599	48	47	45	43	42
AFP275215	82	80	70	2	55
AFP290397	149	148	2	1	29
AFP306591	45	44	84	83	65
AFP310297	23	31	37	47	30
AFP314720	47	44	26	25	23
AFP318671	55	54	51	64	63
AFP323575	75	73	72	70	18
AFP327160	37	68	47	67	96
AFP329002	78	77	76	75	74
AFP345415	41	40	133	106	39
AFP347179	30	4	29	86	177
AFP359138	77	2	76	75	74
AFP365372	13	1	62	69	79
AFP367284	61	60	36	5	59
AFP372822	49	48	25	8	24
AFP374595	154	153	165	3	56
AFP375952	36	35	53	52	69
AFP382913	67	32	30	20	66
AFP389184	24	31	78	30	39
AFP404208	69	68	67	39	36
AFP404279	81	31	72	30	62

AFP409112	97	96	56	94	55
AFP413111	65	85	96	64	94
AFP415635	35	26	25	34	32
AFP421092	27	1	46	57	35
AFP436666	5	95	59	4	58
AFP448623	14				
AFP454192	106	104	83	114	112
AFP49026	49	104	76	48	138
AFP51688	51	86	50	85	43
AFP525341	18	17	16	79	14
AFP545268	65	64	75	21	74
AFP592620	22	21	29	20	28
AFP62197	134	84	133	20	104
AFP68229	161	171	192	170	232
AFP71288	67	49	65	48	46
AFP77851	123	121	33	103	53
AFP81957	89	66	63	25	40
AFP85168	61	31	39	27	46

As used herein, the term "antibodies" includes polyclonal antibodies, monoclonal antibodies, antigen-binding fragments thereof such as F(ab')<sub>2</sub> and Fab fragments, single chain antibodies, and the like, including genetically engineered antibodies. Non-human antibodies can be humanized by grafting only non-human CDRs onto human framework and constant regions, or by incorporating the entire non-human variable domains (optionally "cloaking" them with a human-like surface by replacement of exposed residues, wherein the result is a "veneered" antibody). In some instances, humanized antibodies may retain non-human residues within the human variable region framework domains to enhance proper binding characteristics. Through humanizing antibodies, biological half-life may be increased, and the potential for adverse immune reactions upon administration to humans is reduced. One skilled in the art can generate humanized antibodies with specific and different constant domains (i.e., different Ig subclasses) to facilitate or inhibit various immune functions associated with particular antibody constant domains.

Alternative techniques for generating or selecting antibodies useful herein include *in vitro* exposure of lymphocytes to an immunogenic polypeptide, and selection of antibody display libraries in phage or similar vectors (for instance, through use of an immobilized or labeled polypeptide). Human antibodies can be produced in

transgenic, non-human animals that have been engineered to contain human immunoglobulin genes as disclosed in WIPO Publication WO 98/24893. It is preferred that the endogenous immunoglobulin genes in these animals be inactivated or eliminated, such as by homologous recombination.

5           Antibodies are defined to be specifically binding if they bind to a target polypeptide with an affinity at least 10-fold greater than the binding affinity to control (non-target) polypeptide. It is preferred that the antibodies exhibit a binding affinity ( $K_d$ ) of  $10^6 M^{-1}$  or greater, preferably  $10^7 M^{-1}$  or greater, more preferably  $10^8 M^{-1}$  or greater, and most preferably  $10^9 M^{-1}$  or greater. The affinity of a monoclonal antibody  
10 can be readily determined by one of ordinary skill in the art (see, for example, Scatchard, *Ann. NY Acad. Sci.* 51: 660-672, 1949).

          Methods for preparing polyclonal and monoclonal antibodies are well known in the art (see for example, Hurrell, J. G. R., Ed., *Monoclonal Hybridoma Antibodies: Techniques and Applications*, CRC Press, Inc., Boca Raton, FL, 1982). As  
15 would be evident to one of ordinary skill in the art, polyclonal antibodies can be generated from a variety of warm-blooded animals such as horses, cows, goats, sheep, dogs, chickens, rabbits, mice, and rats. The immunogenicity of a polypeptide immunogen may be increased through the use of an adjuvant such as alum (aluminum hydroxide) or Freund's complete or incomplete adjuvant. Polypeptides useful for  
20 immunization also include fusion polypeptides, such as fusions of a polypeptide of interest or a portion thereof with an immunoglobulin polypeptide or with maltose binding protein. The polypeptide immunogen may be a full-length molecule or a portion thereof. If the polypeptide portion is "haptene-like", such portion may be advantageously joined or linked to a macromolecular carrier (such as keyhole limpet  
25 hemocyanin (KLH), bovine serum albumin (BSA) or tetanus toxoid) for immunization.

          A variety of assays known to those skilled in the art can be utilized to detect antibodies that specifically bind to a polypeptide of interest. Exemplary assays are described in detail in *Antibodies: A Laboratory Manual*, Harlow and Lane (Eds.), Cold Spring Harbor Laboratory Press, 1988. Representative examples of such assays  
30 include concurrent immunoelectrophoresis, radio-immunoassays, radio-immunoprecipitations, enzyme-linked immunosorbent assays (ELISA), dot blot assays, Western blot assays, inhibition or competition assays, and sandwich assays.

          Antibodies can be used, for example, to isolate target polypeptides by affinity purification, for diagnostic assays for determining circulating or localized  
35 levels of target polypeptides, for tissue typing, for cell sorting, for screening expression libraries; for generating anti-idiotypic antibodies, and as neutralizing antibodies or as antagonists to block protein activity *in vitro* and *in vivo*.

The present invention also provides reagents for use in diagnostic and therapeutic applications. Such reagents include polynucleotide probes and primers; antibodies, including antibody fragments, single-chain antibodies, and other genetically engineered forms; soluble receptors and other polypeptide binding partners; and the proteins of the invention themselves, including fragments thereof. Those skilled in the art will recognize that diagnostic reagents will commonly be labeled to provide a detectable signal or other second function. Thus, polypeptides, antibodies, receptors, and other binding partners disclosed herein can be directly or indirectly conjugated to drugs, toxins, radionuclides, enzymes, enzyme substrates, cofactors, inhibitors, fluorescent markers, chemiluminescent markers, magnetic particles, and the like, and these conjugates used for *in vivo* diagnostic or therapeutic applications. Cytotoxic molecules, for example, can be directly or indirectly attached to the binding partner (e.g., by chemical coupling or as a fusion protein), and include bacterial or plant toxins (e.g., diphtheria toxin, *Pseudomonas* exotoxin, ricin, saporin, abrin, and the like); therapeutic radionuclides (e.g., iodine-131, rhenium-188 or yttrium-90) which can be directly attached to a polypeptide or antibody or indirectly attached through means of a chelating moiety; and cytotoxic drugs (e.g., adriamycin). Methods for preparing labeled reagents are known in the art. Within an alternative embodiment, the detectable signal or other function can be provided by a second member of a complement-anticomplement pair, which second member binds to the diagnostic reagent. For example, a first (unlabeled) antibody can be used to bind to a cell-surface polypeptide, after which a second, labeled antibody which binds to the first antibody is added. Other complement-anticomplement pairs are known in the art and include biotin/streptavidin.

Diagnostic reagents as disclosed herein can be used *in vivo* or *in vitro*. *In vitro* diagnostic assays include assays of tissue and fluid samples. Assays for protein in serum, for example, may be used to detect metabolic abnormalities characterized by over- or under-production of the protein, such as cancers, immune system abnormalities, infections, organ failure, metabolic imbalances, inborn errors of metabolism and other disease states. Proteins of the present invention can also be used in the detection of circulating autoantibodies, which are indicative of autoimmune disorders. Those skilled in the art will recognize that conditions related to protein underexpression or overexpression may be amenable to treatment by therapeutic manipulation of the relevant protein level(s). Proteins in serum can be quantitated by known methods known in the art, which include the use of antibodies in a variety of formats. Non-antibody binding partners, such as ligand-binding receptor fragments (commonly referred to as "soluble receptors") can also be used.

In general, diagnostic methods employing oligonucleotide probes or primers comprise the steps of (a) obtaining a genetic sample from a patient; (b) incubating the genetic sample with an oligonucleotide probe or primer as disclosed above, under conditions wherein the probe or primer will hybridize to a complementary polynucleotide sequence, to produce a first reaction product; and (c) comparing the first reaction product to a control reaction product. A difference between the first reaction product and the control reaction product is indicative of a genetic abnormality in the patient. Genetic samples for use within such methods include genomic DNA, cDNA, and RNA. Suitable assay methods in this regard include molecular genetic techniques known to those in the art, such as restriction fragment length polymorphism (RFLP) analysis, short tandem repeat (STR) analysis employing PCR techniques, ligation chain reaction (Barany, *PCR Methods and Applications* 1:5-16, 1991), ribonuclease protection assays, and other genetic linkage analysis techniques known in the art (Sambrook et al., *ibid.*; Ausubel et al., *ibid.*; A.J. Marian, *Chest* 108:255-65, 1995). Ribonuclease protection assays (see, e.g., Ausubel et al., *ibid.*, ch. 4) comprise the hybridization of an RNA probe to a patient RNA sample, after which the reaction product (RNA-RNA hybrid) is exposed to RNase. Hybridized regions of the RNA are protected from digestion. Within PCR assays, a patient genetic sample is incubated with a pair of oligonucleotide primers, and the region between the primers is amplified and recovered. Changes in size, amount, or sequence of recovered product are indicative of mutations in the patient. Another PCR-based technique that can be employed is single strand conformational polymorphism (SSCP) analysis (Hayashi, *PCR Methods and Applications* 1:34-38, 1991). Chromosomal localization data can be used to correlate AFP gene locations with known genetic disorders using, for example, the OMIM™ Database, Johns Hopkins University, 2000 (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>).

Relative chromosomal sublocalization shown in Table 11 was determined using the Draft Human Genome Browser (Kent, J., University of California Santa Cruz, <http://genome.ucsc.edu/goldenPath/hgTracks.html>) displaying the draft assembly of the July 17, 2000 version of the human genome. Table 11 also correlates AFP sequences with corresponding sequences in public databases by GenBank Accession Number, source clone ID number, and EST accession number. Also see Table 5, above.

Table 11

AFP	GenBank Acc. No.	Source Clone ID No.	EST Acc. No.	Chr.	Band	Start	Stop
AFP127023	AP001155	RP11-594B10	*	18	18q12	35729370	35952786
AFP138504	AP001931	RP11-691N7	*	11	11p11.11	53438038	53888802
AFP138740	AC024059	RP11-79121	AW580814	15	15q22.1	58185489	58481462
AFP138740	*	*	AW580814	15		58258653	58308652
AFP177000	AL118506	RP4-591C20	*	20	20q12	48950838	49160243
AFP178828	AC007686	CTD-2289B16;RP11-116N21;RP11-7F17	*	14	14q23.3	62132030	62313415
AFP179530	AC011475	CTC-539A10	*	12	12q12	41234876	41456630
AFP188135	AC013740	*	*	9	9q31.2	91150313	91361876
AFP194554	AC024888	RP11-901L	*	16	16q22.1	71944378	72167142
AFP199044	AC012180	RP11-31110	*	16	16q11.2	44574019	44904017
AFP199200	CNS01DV7	BAC-R-1070N10	*	14		82330266	82541053
AFP229269	AL161670	BAC-R-804M7	*	14	14q21.3	46135365	46299284
AFP236718	AC010319	CTD-2521M24	*	19	19p13.3	4839920	5087628
AFP237679	Z69709	*	*	4	4p16.3	4521455	4544888
AFP244615	*	*	AI494556;AW85055 3	3	3q13.12	11646693	116517043
AFP249599	AL157714	RP11-541H12	*	1	1q22-23.3	161893354	162136704
AFP250422	AC012046	RP11-312P12	*	10	10q22.1	81289799	81650062
AFP262739	AC005884	hRPK.264_B_14	*	17	17q23.3	64245127	64365313
AFP275580	AC016773	*	*	3	3q21.3	141329005	141513510
AFP277451	AC055822	RP11-707M3	*	8	8q13.3	75395740	75583383
AFP279267	*	*	AI566086	10	10q11.1	52859924	52861338
AFP280451	AL133355	RP11-541N10	*	10	10q24.32	115276306	115467187
AFP290397	*	*	AA421069	15	15q15.3	48427462	48427830
AFP293220	AC012476	RP11-532F12	*	15	15p11.1	17263661	17480097
AFP297548	*	*	W52728	11	11q11	57918740	57927327
AFP306591	AQ079258	2366B9	AW118928	6	6p22.3	19812023	19812791
AFP313600	AC005037	NH0469M07	*	2	2q33.1	205320800	205511307
AFP324816	AC011687	RP11-15120	*	2	2p21	49054619	49249783
AFP325761	AC012485	RP11-5024	*	2	2p24.3	17554756	17765537



AFP326051	AL132639	BAC-R-407N17	AI525611	14	14p11.1	19959493	20153358
AFP345861	AC015936	CTD-2534I21	*	17	17q21.2	44087441	44286594
AFP347179	AC025740	*	*	12	12q24.23	125918909	126134148
AFP372822	AL022240	3.28E+21	*	1	1q12-21.2	138667522	138765140
AFP374312	*	*	AI253088	11	11q23.3	128134250	128134589
AFP375952	*	*	AI741157	16	16p13.3	3479999	3500834
AFP395942	AC004235	*	*	16	16p13.3	4189155	4222465
AFP404202	*	*	AI133727	7		142961410	143641730
AFP404279	*	*	AI341602	4	4p16.3	1512179	1514256
AFP413680	AC006942	cosmid-R31181	*	19	19q13.33	59897688	59940397
AFP436666	*	*	AI814257	8	8p21.3	18993217	19003942
AFP448623	*	*	AI140615	5	5q33.1	173540737	173547400
AFP460626	AC009131	RP11-502K10	*	16	16q22.1	70222075	70471703
AFP477303	AC008686	CTB-5E10	*	19	19p13.13	16491516	16677574
AFP501809	*	*	AW583171	6	6p21.1	50554924	50564907
AFP545268	AL138695	RP11-342J4	*	13	13q21.1	60450247	60714738
AFP561930	AL136221	RP11-391H12	*	13	13q34	108494503	108794286
AFP71288	*	*	AA493506	6	6q22.33	137477811	137478427
AFP74517	HS1056L3	RP5-1056L3	*	1	1p35.1-36.13	*	*
AFP93743	AC067942	RP11-791G16	*	4	4q21.22	77419530	77633569

If a mammal has an insufficiency of a protein of interest (due to, for example, a mutated or absent gene), the corresponding wild-type gene can be introduced into the cells of the mammal. In one embodiment, a gene encoding a protein of interest is introduced into the animal using a viral vector. Such vectors include an attenuated or defective DNA virus, such as, but not limited to, herpes simplex virus (HSV), papillomavirus, Epstein Barr virus (EBV), adenovirus, adeno-associated virus (AAV), and the like. Defective viruses, which entirely or almost entirely lack viral genes, are preferred. A defective virus is not infective after introduction into a cell. Use of defective viral vectors allows for administration to cells in a specific, localized area, without concern that the vector can infect other cells. Examples of particular vectors include, but are not limited to, a defective herpes simplex virus 1 (HSV1) vector (Kaplitt et al., *Molec. Cell. Neurosci.* 2:320-30, 1991); an attenuated adenovirus vector, such as the vector described by Stratford-Perricaudet et al. (*J. Clin. Invest.* 90:626-30, 1992); and a defective adeno-associated virus vector (Samulski et al., *J. Virol.* 61:3096-101, 1987; Samulski et al., *J. Virol.* 63:3822-28, 1989).

Within another embodiment, a gene of interest is introduced into an animal by liposome-mediated transfection ("lipofection") essentially as disclosed above. Lipofection can be used to introduce exogenous genes into specific organs.

A gene of interest can also be introduced into an animal for gene therapy as a naked DNA plasmid using the methods disclosed above.

In another embodiment, polypeptide-toxin fusion proteins or antibody/fragment-toxin fusion proteins may be used for targeted cell or tissue inhibition or ablation, such as in cancer therapy. Of particular interest in this regard are conjugates of an AFP protein and a cytotoxin, which can be used to target the cytotoxin to a tumor or other tissue that is undergoing undesired angiogenesis or neovascularization.

In another embodiment, AFP-cytokine fusion proteins or antibody/fragment-cytokine fusion proteins may be used for enhancing *in vitro* cytotoxicity (for instance, that mediated by monoclonal antibodies against tumor targets) and for enhancing *in vivo* killing of target tissues (for example, blood and bone marrow cancers). See, generally, Hornick et al., *Blood* 89:4437-4447, 1997). In general, cytokines are toxic if administered systemically. The described fusion proteins enable targeting of a cytokine to a desired site of action, such as a cell having binding sites for an AFP protein, thereby providing an elevated local concentration of cytokine. Polypeptides, antibodies, or receptors target an undesirable cell or tissue

(e.g., a tumor), and the fused cytokine mediates improved target cell lysis by effector cells. Suitable cytokines for this purpose include, for example, interleukin-2 and granulocyte-macrophage colony-stimulating factor (GM-CSF).

In another embodiment, polypeptide-toxin fusion proteins or other  
5 binding partner-linked toxins may be used for targeted cell or tissue inhibition or ablation (for instance, to treat cancer cells or tissues). Target cells (i.e., those displaying a receptor for a polypeptide of interest) bind the polypeptide-toxin conjugate, which is then internalized, killing the cell. The effects of receptor-specific cell killing (target ablation) are revealed by changes in whole animal physiology or  
10 through histological examination. Thus, ligand-dependent, receptor-directed cytotoxicity can be used to enhance understanding of the physiological significance of a protein ligand. A preferred such toxin is saporin. Mammalian cells have no receptor for saporin, which is non-toxic when it remains extracellular. Alternatively, if the polypeptide of interest has multiple functional domains (i.e., an activation domain or a  
15 ligand binding domain, plus a targeting domain), a fusion protein including only the targeting domain may be suitable for directing a detectable molecule, a cytotoxic molecule or a complementary molecule to a cell or tissue type of interest. In instances where the domain-only fusion protein includes a complementary molecule, the anti-complementary molecule can be conjugated to a detectable or cytotoxic molecule.  
20 Such domain-complementary molecule fusion proteins thus represent a generic targeting vehicle for cell- or tissue-specific delivery of generic anti-complementary-detectable/cytotoxic molecule conjugates.

The bioactive conjugates described herein can be delivered intravenously, intraarterially or intraductally, or may be introduced locally at the  
25 intended site of action.

For pharmaceutical use, the proteins of the present invention are formulated according to conventional methods. Routes of delivery include topical, mucosal, and parenteral, the latter including intravenous and subcutaneous delivery. Intravenous administration will be by bolus injection or infusion over a typical period  
30 of one to several hours. In general, pharmaceutical formulations will include a protein of the present invention in combination with a pharmaceutically acceptable vehicle, such as saline, buffered saline, 5% dextrose in water or the like. Formulations may further include one or more excipients, diluents, fillers, emulsifiers, preservatives, solubilizers, buffering agents, wetting agents, stabilizers, colorings, penetration  
35 enhancers, albumin to prevent protein loss on vial surfaces, etc. Topical formulations are typically provided as liquids, ointments, salves, gels, emulsions and the like. Methods of formulation are well known in the art and are disclosed, for example, in

*Remington: The Science and Practice of Pharmacy*, Gennaro, ed., Mack Publishing Co., Easton, PA, 19th ed., 1995. Therapeutic doses will be determined by the clinician according to accepted standards, taking into account the nature and severity of the condition to be treated, patient traits, etc. Proteins of the present invention will  
5 generally be formulated to provide a dose of from 0.01  $\mu$ g to 100 mg per kg patient weight per day, more commonly from 0.1  $\mu$ g to 10 mg/kg/day, still more commonly from 0.1  $\mu$ g to 1.0 mg/kg/day. Determination of dose is within the level of ordinary skill in the art. The proteins may be administered for acute treatment, over one week or less, often over a period of one to three days or may be used in chronic treatment, over  
10 several months or years. In general, a therapeutically effective amount is an amount sufficient to produce a clinically significant change in the targetted condition.

Within the laboratory research field, the proteins of the present invention can be used as molecular weight standards, or as standards in the analysis of cell phenotype, and as reagents for the study of cells, receptors, and other binding  
15 molecules. Such reagents will generally further comprise a second moiety, such as a label, binding partner, or toxin, that facilitates the detection of the protein when bound to its target. Many such systems are known in the art and are summarized above. Receptors and other cell-surface binding sites for proteins of the present invention can be identified by exposing a population of cells to a labelled protein under physiologic  
20 conditions, whereby the protein binds to the surface of the cell. Cells bearing receptors for a protein of interest can also be identified using the protein joined to a toxin, whereby receptor-bearing cells are killed by the toxin.

AFP proteins and antagonists thereof can be used as standards in assays of protein and protein inhibitors in both clinical and research settings. Such assays can  
25 comprise any of a number of standard formats, include radioreceptor assays and ELISAs. Protein standards can be prepared in labeled form using a radioisotope, enzyme, fluorophore, or other compound that produces a detectable signal. The proteins can be packaged in kit form, such kits comprising one or more vials containing the AFP protein and, optionally, a diluent, an antibody, a labeled binding protein, etc.  
30 Assay kits can be used in the research laboratory to detect protein and inhibitor activities produced by cultured cells or test animals.

Proteins of the present invention may also be used as protein and amino acid supplements, including hydrolysates. Specific uses in this regard include use as animal feed supplements and as cell culture components. Proteins rich in a particular  
35 amino acid can be used as a source of that amino acid.

Polynucleotides and polypeptides of the present invention will additionally find use as educational tools as a laboratory practicum kits for courses

related to genetics and molecular biology, protein chemistry and antibody production and analysis. Due to their unique polynucleotide and polypeptide sequences, molecules of AFP protein or polynucleotide can be used as standards or as "unknowns" for testing purposes. For example, AFP polynucleotides can be used as aids in  
5 teaching students how to prepare expression constructs for bacterial, viral, and/or mammalian expression, including fusion constructs, wherein an AFP polynucleotide is the gene to be expressed; for determining the restriction endonuclease cleavage sites of the polynucleotides (which can be determined from the sequence using conventional computer software, such as MapDraw™ (DNASTAR, Madison, WI)); determining  
10 mRNA and DNA localization of AFP polynucleotides in tissues (e.g., by Northern and Southern blotting as well as polymerase chain reaction); and for identifying related polynucleotides and polypeptides by nucleic acid hybridization.

AFP polypeptides can be used educationally as aids to teach preparation of antibodies; identifying proteins by Western blotting; protein purification;  
15 determining the weight of expressed AFP polypeptides as a ratio to total protein expressed; identifying peptide cleavage sites; coupling amino and carboxyl terminal tags; amino acid sequence analysis, as well as, but not limited to monitoring biological activities of both the native and tagged protein (i.e., receptor binding, signal transduction, proliferation, and differentiation) *in vitro* and *in vivo*. AFP polypeptides  
20 can also be used to teach analytical skills such as mass spectrometry, circular dichroism to determine conformation, in particular the locations of the disulfide bonds, x-ray crystallography to determine the three-dimensional structure in atomic detail, nuclear magnetic resonance spectroscopy to reveal the structure of proteins in solution. For example, a kit containing an AFP protein can be given to the student to analyze.  
25 Since the amino acid sequence would be known by the professor, the protein can be given to the student as a test to determine the skills or develop the skills of the student, the teacher would then know whether or not the student has correctly analyzed the polypeptide. Since every polypeptide is unique, the educational utility of zcub5 would be unique unto itself.

30 Antibodies that bind specifically to an AFP polypeptide can be used as a teaching aid to instruct students how to prepare affinity chromatography columns to purify the cognate polypeptide, cloning and sequencing the polynucleotide that encodes an antibody and thus as a practicum for teaching a student how to design humanized antibodies. The AFP polynucleotide, polypeptide or antibody would then be packaged  
35 by reagent companies and sold to universities so that the students gain skill in art of molecular biology. Because each polynucleotide and protein is unique, each polynucleotide and protein creates unique challenges and learning experiences for

students in a lab practicum. Such educational kits containing an AFP polynucleotide, polypeptide or antibody are considered within the scope of the present invention.

The invention is further illustrated by the following non-limiting examples.

5

### EXAMPLES

#### Example 1

A protein of the present invention ("AFP") is produced in *E. coli* using a His<sub>6</sub> tag/maltose binding protein (MBP) double affinity fusion system as generally disclosed by Pryor and Leiting, *Prot. Expr. Pur.* 10:309-319, 1997. A thrombin cleavage site is placed at the junction between the affinity tag and AFP sequences.

The fusion construct is assembled in the vector pTAP98, which comprises sequences for replication and selection in *E. coli* and yeast, the *E. coli* tac promoter, and a unique SmaI site just downstream of the MBP-His<sub>6</sub>-thrombin site coding sequences. The AFP cDNA is amplified by PCR using primers each comprising 40 bp of sequence homologous to vector sequence and 25 bp of sequence that anneals to the cDNA. The reaction is run using Taq DNA polymerase (Boehringer Mannheim, Indianapolis, IN) for 30 cycles of 94°C, 30 seconds; 60°C, 60 seconds; and 72°C, 60 seconds. One microgram of the resulting fragment is mixed with 100 ng of SmaI-cut pTAP98, and the mixture is transformed into yeast to assemble the vector by homologous recombination (Oldenburg et al., *Nucl. Acids. Res.* 25:451-452, 1997). Ura<sup>+</sup> transformants are selected.

Plasmid DNA is prepared from yeast transformants and transformed into *E. coli* MC1061. Pooled plasmid DNA is then prepared from the MC1061 transformants by the miniprep method after scraping an entire plate. Plasmid DNA is analyzed by restriction digestion.

*E. coli* strain BL21 is used for expression of AFP. Cells are transformed by electroporation and grown on minimal glucose plates containing casamino acids and ampicillin.

Protein expression is analyzed by gel electrophoresis. Cells are grown in liquid glucose media containing casamino acids and ampicillin. After one hour at 37°C, IPTG is added to a final concentration of 1mM, and the cells are grown for an additional 2-3 hours at 37°C. Cells are disrupted using glass beads, and extracts are prepared.

35

Example 2

Larger scale cultures of AFP transformants are prepared by the method of Pryor and Leiting (*ibid.*). 100-ml cultures in minimal glucose media containing casamino acids and 100 µg/ml ampicillin are grown at 37°C in 500-ml baffled flasks to  
5 OD<sub>600</sub> ≈ 0.5. Cells are harvested by centrifugation and resuspended in 100 ml of the same media at room temperature. After 15 minutes, IPTG is added to 0.5 mM, and cultures are incubated at room temperature (ca. 22.5°C) for 16 to 20 hours with shaking at 125 rpm. The culture is harvested by centrifugation, and cell pellets are stored at -70°C.

10

Example 3

For larger-scale protein preparation, 500-ml cultures of *E. coli* BL21 expressing the AFP-MBP-His<sub>6</sub> fusion protein are prepared essentially as disclosed in Example 2. Cell pellets are resuspended in 100 ml of binding buffer (20 mM Tris, pH  
15 7.58, 100 mM NaCl, 20 mM NaH<sub>2</sub>PO<sub>4</sub>, 0.4 mM 4-(2-Aminoethyl)-benzenesulfonyl fluoride hydrochloride [Pefabloc® SC; Boehringer-Mannheim], 2 µg/ml Leupeptin, 2 µg/ml Aprotinin). The cells are lysed in a French press at 30,000 psi, and the lysate is centrifuged at 18,000 x g for 45 minutes at 4°C to clarify it. Protein concentration is estimated by gel electrophoresis with a BSA standard.

20 Recombinant AFP fusion protein is purified from the lysate by affinity chromatography. Immobilized cobalt resin (Talon® resin; Clontech Laboratories, Inc., Palo Alto, CA) is equilibrated in binding buffer. One ml of packed resin per 50 mg protein is combined with the clarified supernatant in a tube, and the tube is capped and sealed, then placed on a rocker overnight at 4°C. The resin is then pelleted by  
25 centrifugation at 4°C and washed three times with binding buffer. Protein is eluted with binding buffer containing 0.2 M imidazole. The resin and elution buffer are mixed for at least one hour at 4°C, the resin is pelleted, and the supernatant is removed. An aliquot is analyzed by gel electrophoresis, and concentration is estimated. Amylose resin is equilibrated in amylose binding buffer (20 mM Tris-HCl, pH 7.0, 100 mM  
30 NaCl, 10 mM EDTA) and combined with the supernatant from the Talon resin at a ratio of 2 mg fusion protein per ml of resin. Binding and washing steps are carried out as disclosed above. Protein is eluted with amylose binding buffer containing 10 mM maltose using as small a volume as possible to minimize the need for subsequent concentration. The eluted protein is analyzed by gel electrophoresis and staining with  
35 Coomassie blue using a BSA standard, and by Western blotting using an anti-MBP antibody.

#### Example 4

An expression plasmid containing all or part of a polynucleotide encoding AFP is constructed via homologous recombination. An AFP coding sequence comprising the ORF with 5' and 3' ends corresponding to the vector sequences flanking the insertion point is prepared by PCR. The primers for PCR each include from 5' to 3' end: 40 bp of flanking sequence from the vector and 17 bp corresponding to the amino or carboxyl termini from the open reading frame of AFP.

Ten  $\mu$ l of the 100  $\mu$ l PCR reaction mixture is run on a 0.8% low-melting-temperature agarose (SeaPlaque GTG®; FMC BioProducts, Rockland, ME) gel with 1 x TBE buffer for analysis. The remaining 90  $\mu$ l of the reaction mixture is precipitated with the addition of 5  $\mu$ l 1 M NaCl and 250  $\mu$ l of absolute ethanol. The plasmid pZMP6, which has been cut with SmaI, is used for recombination with the PCR fragment. Plasmid pZMP6 is a mammalian expression vector containing an expression cassette having the cytomegalovirus immediate early promoter, multiple restriction sites for insertion of coding sequences, a stop codon, and a human growth hormone terminator; an *E. coli* origin of replication; a mammalian selectable marker expression unit comprising an SV40 promoter, enhancer and origin of replication, a DHFR gene, and the SV40 terminator; and URA3 and CEN-ARS sequences required for selection and replication in *S. cerevisiae*. It was constructed from pZP9 (deposited at the American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209, under Accession No. 98668) with the yeast genetic elements taken from pRS316 (available from the American Type Culture Collection, 10801 University Boulevard, Manassas, VA, under Accession No. 77145), an internal ribosome entry site (IRES) element from poliovirus, and the extracellular domain of CD8 truncated at the C-terminal end of the transmembrane domain.

One hundred microliters of competent yeast (*S. cerevisiae*) cells are independently combined with 10  $\mu$ l of the various DNA mixtures from above and transferred to a 0.2-cm electroporation cuvette. The yeast/DNA mixtures are electropulsed using power supply (BioRad Laboratories, Hercules, CA) settings of 0.75 kV (5 kV/cm),  $\infty$  ohms, 25  $\mu$ F. To each cuvette is added 600  $\mu$ l of 1.2 M sorbitol, and the yeast is plated in two 300- $\mu$ l aliquots onto two URA-D plates (1.8% agar in 2% D-glucose, 0.67% yeast nitrogen base without amino acids, 0.056% -Ura -Trp -Thr powder [made by combining 4.0 g L-adenine, 3.0 g L-arginine, 5.0 g L-aspartic acid, 2.0 g L-histidine, 6.0 g L-isoleucine, 8.0 g L-leucine, 4.0 g L-lysine, 2.0 g L-methionine, 6.0 g L-phenylalanine, 5.0 g L-serine, 5.0 g L-tyrosine, and 6.0 g L-valine], and 0.5% 200X tryptophan, threonine solution [3.0% L-threonine, 0.8% L-tryptophan in H<sub>2</sub>O]) and incubated at 30°C. After about 48 hours, the Ura<sup>+</sup> yeast



transformants from a single plate are resuspended in 1 ml H<sub>2</sub>O and spun briefly to pellet the yeast cells. The cell pellet is resuspended in 1 ml of lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris, pH 8.0, 1 mM EDTA). Five hundred microliters of the lysis mixture is added to an Eppendorf tube containing 300 µl acid-washed glass beads and 200 µl phenol-chloroform, vortexed for 1 minute intervals two or three times, and spun for 5 minutes in an Eppendorf centrifuge at maximum speed. Three hundred microliters of the aqueous phase is transferred to a fresh tube, and the DNA is precipitated with 600 µl ethanol (EtOH), followed by centrifugation for 10 minutes at 4°C. The DNA pellet is resuspended in 10 µl H<sub>2</sub>O.

Transformation of electrocompetent *E. coli* host cells (Electromax DH10B™ cells; obtained from Life Technologies, Inc., Gaithersburg, MD) is done with 0.5-2 ml yeast DNA prep and 40 µl of cells. The cells are electropulsed at 1.7 kV, 25 µF, and 400 ohms. Following electroporation, 1 ml SOC (2% Bacto™ Tryptone (Difco, Detroit, MI), 0.5% yeast extract (Difco), 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub>, 20 mM glucose) is plated in 250-µl aliquots on four LB AMP plates (LB broth (Lennox), 1.8% Bacto™ Agar (Difco), 100 mg/L Ampicillin).

Individual clones harboring the correct expression construct for AFP are identified by restriction digest to verify the presence of the AFP insert and to confirm that the various DNA sequences have been joined correctly to one another. The inserts of positive clones are subjected to sequence analysis. Larger scale plasmid DNA is isolated using a commercially available kit (QIAGEN Plasmid Maxi Kit, Qiagen, Valencia, CA) according to manufacturer's instructions. The correct construct is designated pZMP6/AFP.

Recombinant protein is produced in BHK cells transfected with pZMP6/AFP. BHK 570 cells (ATCC CRL-10314) are plated in 10-cm tissue culture dishes and allowed to grow to approximately 50 to 70% confluence overnight at 37°C, 5% CO<sub>2</sub>, in DMEM/FBS media (DMEM, Gibco/BRL High Glucose; Life Technologies), 5% fetal bovine serum (Hyclone, Logan, UT), 1 mM L-glutamine (JRH Biosciences, Lenexa, KS), 1 mM sodium pyruvate (Life Technologies). The cells are then transfected with pZMP6/AFP by liposome-mediated transfection using a 3:1 (w/w) liposome formulation of the polycationic lipid 2,3-dioleoyloxy-N-[2(sperminecarboxamido)ethyl]-N,N-dimethyl-1-propaniminium-trifluoroacetate and the neutral lipid dioleoyl phosphatidylethanolamine in membrane-filtered water (Lipofectamine™ Reagent; Life Technologies, Gaithersburg, MD), in serum free (SF) media (DMEM supplemented with 10 mg/ml transferrin, 5 mg/ml insulin, 2 mg/ml fetuin, 1% L-glutamine and 1% sodium pyruvate). The plasmid is diluted into 15-ml tubes to a total final volume of 640 µl with SF media. 35 µl of the lipid mixture is

mixed with 605  $\mu$ l of SF medium, and the resulting mixture is allowed to incubate approximately 30 minutes at room temperature. Five milliliters of SF media is then added to the DNA:lipid mixture. The cells are rinsed once with 5 ml of SF media, aspirated, and the DNA:lipid mixture is added. The cells are incubated at 37°C for five  
5 hours, then 6.4 ml of DMEM/10% FBS, 1% PSN media is added to each plate. The plates are incubated at 37°C overnight, and the DNA:lipid mixture is replaced with fresh 5% FBS/DMEM media the next day. On day 5 post-transfection, the cells are split into T-162 flasks in selection medium (DMEM + 5% FBS, 1% L-Gln, 1% NaPyr, 1  $\mu$ M methotrexate). Approximately 10 days post-transfection, two 150-mm culture  
10 dishes of methotrexate-resistant colonies from each transfection are trypsinized, and the cells are pooled and plated into a T-162 flask and transferred to large-scale culture.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration,  
15 various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

## CLAIMS

We claim:

1. An isolated polypeptide comprising fifteen contiguous amino acid residues of a polypeptide as shown in SEQ ID NO:M, wherein M is an even integer from 2 to 422.
2. The isolated polypeptide of claim 1 wherein M is 6, 8, 12, 18, 24, 42, 48, 54, 66, 68, 70, 72, 90, 92, 96, 98, 102, 106, 110, 122, 134, 138, 140, 156, 158, 162, 164, 168, 174, 178, 180, 204, 206, 210, 224, 230, 234, 236, 240, 242, 252, 254, 258, 270, 272, 284, 286, 288, 294, 300, 302, 306, 312, 314, 324, 326, 338, 342, 344, 348, 350, 366, 368, 374, 378, 386, 388, 396, 398, 402, 408, 412, or 416.
3. The isolated polypeptide of claim 1 or claim 2 which is from 15 to 2235 amino acid residues in length.
4. The isolated polypeptide of claim 3 which is operably linked via a peptide bond or polypeptide linker to a second polypeptide selected from the group consisting of maltose binding protein, an immunoglobulin constant region, a polyhistidine tag, and a peptide as shown in SEQ ID NO:423.
5. The isolated polypeptide of any of claims 1-4 comprising at least 30 contiguous residues of SEQ ID NO:M.
6. The isolated polypeptide of any of claims 1-5 comprising at least 47 contiguous residues of SEQ ID NO:M.
7. An isolated, mature protein encoded by a sequence selected from the group consisting of SEQ ID NO:N, wherein N is an odd integer from 1 to 421.
8. The protein of claim 7 wherein N is 5, 7, 11, 17, 23, 41, 47, 53, 65, 67, 69, 71, 89, 91, 95, 97, 101, 105, 109, 121, 133, 137, 139, 155, 157, 161, 163, 167, 173, 177, 179, 203, 205, 209, 223, 229, 233, 235, 239, 241, 251, 253, 257, 269, 271, 283, 285, 287, 293, 299, 301, 305, 311, 313, 323, 325, 337, 341, 343, 347, 349, 365, 367, 373, 377, 385, 387, 395, 397, 401, 407, 411, or 415.
9. An isolated polynucleotide comprising a sequence of nucleotides as shown in SEQ ID NO:N, wherein N is an odd integer from 1 to 421.

10. The isolated polynucleotide of claim 9 wherein N is 5, 7, 11, 17, 23, 41, 47, 53, 65, 67, 69, 71, 89, 91, 95, 97, 101, 105, 109, 121, 133, 137, 139, 155, 157, 161, 163, 167, 173, 177, 179, 203, 205, 209, 223, 229, 233, 235, 239, 241, 251, 253, 257, 269, 271, 283, 285, 287, 293, 299, 301, 305, 311, 313, 323, 325, 337, 341, 343, 347, 349, 365, 367, 373, 377, 385, 387, 395, 397, 401, 407, 411, or 415.

11. An expression vector comprising the following operably linked elements:

a transcription promoter;

a DNA segment encoding a polypeptide as shown in SEQ ID NO:M, wherein M is an even integer from 2 to 422; and

a transcription terminator.

12. The expression vector of claim 11 wherein M is 6, 8, 12, 18, 24, 42, 48, 54, 66, 68, 70, 72, 90, 92, 96, 98, 102, 106, 110, 122, 134, 138, 140, 156, 158, 162, 164, 168, 174, 178, 180, 204, 206, 210, 224, 230, 234, 236, 240, 242, 252, 254, 258, 270, 272, 284, 286, 288, 294, 300, 302, 306, 312, 314, 324, 326, 338, 342, 344, 348, 350, 366, 368, 374, 378, 386, 388, 396, 398, 402, 408, 412, or 416.

13. A cultured cell comprising the expression vector of claim 11 or claim 12.

14. A method of producing a polypeptide comprising culturing the cell of claim 13 under conditions whereby said sequence of nucleotides is expressed, and recovering said polypeptide.

15. A polypeptide produced by the method of claim 14.

16. An isolated polynucleotide encoding a fusion protein, said protein comprising a secretory peptide selected from the group consisting of secretory peptides shown in SEQ ID NO:M, wherein M is an even integer from 2 to 422, operably linked to a second polypeptide.

17. An expression vector comprising the following operably linked elements:

a transcription promoter;  
a DNA segment encoding a fusion protein, said protein comprising a secretory peptide selected from the group consisting of secretory peptides shown in SEQ ID NO:M, wherein M is an even integer from 2 to 422, operably linked to a second polypeptide; and  
a transcription terminator.

18. A cultured cell comprising the expression vector of claim 17, wherein the cell expresses the DNA segment and produces the encoded fusion protein.

19. A method of producing a protein comprising culturing the cell of claim 18 under conditions whereby said DNA segment is expressed, and recovering said second polypeptide.

20. An antibody that specifically binds to a protein selected from of the group consisting of SEQ ID NO:M, wherein M is an even integer from 2 to 422.